DOCUMENT RESUME

ED 374 148 TM 022 074

AUTHOR	Keyes, Denis William
TITLE	IQ and the Death Penalty: Verifying Mental Retardation.
PUB DATE	Aug 93
NOTE	41p.; Paper presented at the Annual Meeting of the American Psychological Association (101st, Toronto, Ontario, Canada, August 1993).
PUB TYPE	Reports - Research/Technical (143) -- ...hes/Conference Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Adults; Crime; *Criminals; Difficulty Level; *Intelligence Quotient; Intelligence Tests; Males; *Mental Retardation; Qualitative Research; *Simulation; Statistical Studies
IDENTIFIERS	*Capital Punishment; Stanford Binet Intelligence Scale Fourth Edition; Wechsler Adult Intelligence Scale (Revised)

ABSTRACT

Whether or not subjects can simulate mental retardation, a consideration that has implications in criminal cases, was studied using 21 adult Caucasian males between 20 and 30 years of age, largely comprised of students and staff employees of the University of New Mexico. Subjects were asked to give genuine and simulated responses to two major test batteries of intelligence, the Stanford Binet Fourth Edition and the Wechsler Adult Intelligence Scale Revised. Individual subjects did not appear able to simulate mental retardation consistently on both subtests and achieve statistically similar results. Latency times for simulating appeared to be significantly longer than latency times for genuine responding. Findings further suggested that genuine responses will yield similar scores on both subtests. Qualitative data suggest that it is more difficult to simulate mental retardation than to give genuine answers. Implications for participants in the criminal justice system may be significant, particularly for defendants accused of falsifying a test. Fourteen tables present study data. (Contains 145 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 374 148

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

DENIS W. KEYES

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

IQ and the Death Penalty: Verifying Mental Retardation

A paper Presented to the 101st Annual Conference of the
American Psychological Association
Toronto, Ontario, Canada
August, 1993

Denis William Keyes, Ph.D.
School of Education
University of Charleston, South Carolina

BEST COPY AVAILABLE

INTRODUCTION

Accurate determination of intellectual abilities is important for correct diagnosis of examinee functioning, particularly for individuals being examined during adjudication and sentencing procedures (Burr, 1992). Impartial resolution of court cases may hinge upon the presumed accuracy of these psychological examination results (Ziskin, 1981). However, the accuracy of psychological testing may be difficult to determine, particularly when examinees may have motivations for falsification (Rogers, 1988; Anastasi, 1988).

Two primary reasons for obtaining inaccurate readings are: 1) Using invalid tests; and, 2) deliberate examinee simulation. In most psychological testing, it is assumed that individual examinees will perform as accurately as their abilities permit, displaying their highest possible level of intellect in the resulting scores. Furthermore, initial introduction of testing procedures is usually preceded by minor exchanges of information and some relaxed conversation in order to put the examinee at ease. Once rapport is established, most examinees will put forth the effort necessary to obtain reasonably valid results (Anastasi, 1988). Results of such testing are considered to be reliable and valid if they adequately reflect the true abilities of the examinee (Lezak, 1983; Sattler, 1988).

Reliability and Validity of Test Instruments

All adequately standardized tests of intelligence are continually and thoroughly researched in the effort to determine accurate levels of test reliability and validity. A test's reliability refers to the consistency with which a test will achieve similar results over time and across settings. Validity refers to the accuracy with which a test measures what it purports to measure. Both reliability and validity are very important considerations when designing tests, but if it is not valid, even the most reliable test is virtually useless.

There are several forms of test validity. The level of agreement in results found on scores between different testing instruments is a form of concurrent validity. In this instance, concurrent validity refers to similarity of findings on different tests which purport to measure the same or very similar attributes. If standard scores achieved on one test are highly correlated to the scores achieved on another test, the tests are said to have good concurrent validity (Sattler, 1988).

Major tests of intelligence require high standards in validity, in order to assure reasonable levels of accuracy in measurement. The validity standards for both the Stanford-Binet: Fourth Edition (SB:FE) (Thorndike, Hagen, and Sattler, 1986a) and the Wechsler Adult Intelligence Scale - Revised (WAIS-R) (Wechsler, 1981) indicate that these two tests batteries are well constructed, and that they both are measuring similar psychological attributes, i.e., intelligence. For example, Thorndike, Hagen, and Sattler (1986b) report a correlation of .91 between the Full Scale Intelligence Quotient (IQ) of the WAIS-R and the Composite Score of the SB:FE. This suggests that the psychological attributes measured by each test are nearly identical in nature, indicating acceptable levels of concurrent validity. Breaking this down even further, several specific subtests on both intelligence tests also measure similar or identical attributes (Anastasi, 1988; Sattler, 1988). While subtest content differs widely on each test battery, certain subtests are almost identical in nature. In fact, two subtests from each battery have identical names, Vocabulary and Comprehension.

When raw scores are converted to standard scores, subtest profiles of both the WAIS-R and the SB:FE identify strengths and weaknesses in specific areas of intellectual functioning. Given the properties of concurrent validity, the profiled results of corresponding norm-referenced subtest scores should be very nearly identical. For example, if scores on the vocabulary subtest in the SB:FE indicate a weakness, scores on the same subtest from the WAIS-R should also be lower than average, thus indicating test agreement with respect to an examinee's ability to use vocabulary.

However, if an individual was to deliberately falsify answers to subtest items on one or both tests, a question arises regarding whether or not test validity standards would accurately identify the deception. That is, would an examinee's obtained subtest score on one measure correspond to a score on a separate subtest of similar content, if the examinee was simulating responses? While examinees may not find it particularly difficult to consistently simulate responses, the scoring procedures and standard score calibration of each test battery are very different, and to consistently falsify the same level of ability would be highly unlikely, if not impossible. If the tests truly hold adequate levels of concurrent (criterion-related) validity, the scored false responses should not consistently reflect identical profiles, thus suggesting the possibility or probability of deception. Accordingly, it is doubtful whether examinees could deliberately falsify their responses on both tests and achieve scores which are not, at least statistically, different.

Response Falsification

There may exist occasions where people deliberately falsify their responses to items and tasks administered in order to present an inaccurate profile of their abilities. According to the Diagnostic and Statistical Manual of Mental Disorders - Third Edition (Revised) (DSM III-R) (American Psychiatric Association, 1987), falsification of responses may occur for several reasons. However, the most likely reason would appear to be based upon the examinee's perception that there might be some form of personal gain by performing poorly on the test (Anastasi, 1988; Lezak, 1983). Gain, in this sense, could be operationalized in several ways, including financial, emotional, or legal rewards or concessions obtained from another party.

Sattler (1988) cautions that examiners must be vigilant in determining the level of cooperation and effort exhibited by examinees. Operationalization of these behaviors often presents problems, but they must ultimately be determined based upon the examiner's expertise and clinical awareness. While children may perform at a lower level due to various environmental or emotional factors (Sattler, 1988), adults may perform poorly for other reasons, especially if there is a perceived opportunity for some form of pay-off (Lezak, 1983). The act of malingering, or simulating a physical or psychological condition for purposes of personal gain, presents considerable problems for the clinician.

The Criminal Justice System

Feigning mental illness in order to trick the criminal justice system is not a new phenomenon. Defendants determined to be not guilty by reason of insanity may receive alternative or less severe sentences based upon proof of their mental incapacity. For example, in United States v. Hinckley (525 F. Supp. 1342, 1981), the defendant was found not guilty by reason of insanity. Therefore, due to mental illness, a man accused of attempting to assassinate the President of the United States was deemed to be not culpable for the crime, despite clear and overwhelming evidence of his having committed it. Instead of going to prison, he was committed to a mental hospital. Thus, it is not difficult to understand why criminal defendants may view an insanity defense as one that might well reduce their chances of severe penalties (Grisso, 1986b).

Use of the 'insanity plea' has been in existence for centuries, whereby people are found not guilty of crimes because they lacked the mental or emotional control to refrain from committing them (Grisso, 1986b). While mental illness has been viewed thus, mental retardation has not always been considered a factor for serious consideration of legal implications until recently (Everington. 1987), despite the American Bar Association's Standard including mental retardation as a factor of the mental nonresponsibility defense (ABA Criminal Justice Mental Health Standards, 1989).

Within the criminal justice system, performance on tests of intelligence may constitute important evidence. Lawyers for several national organizations and human rights associations have recently been active in challenging the imposition of the death penalty for convictions where the defendant has mental deficiency and/or neurological impairment. In these situations, evidence of mental retardation must be admitted (Penry v. Lynaugh, 1989). Such evidence may introduce the possibility not only of reduced competence to stand trial (Everington, 1987; Everington & Luckasson, 1990; Everington & Luckasson, 1992), but also of a reduced or specialized sentence (Burr, 1992; Penry v. Lynaugh, 1989).

While intelligence tests clearly do not measure all skill areas, they do provide an important sampling of an individual's abilities in several skill areas related to intelligence (Anastasi, 1988; Jensen, 1980; Kaufman & Kaufman, 1983; Sattler, 1988). When results of a single testing of these abilities are questionable, administration of another test may be desirable as a means of verification. When one considers the proven relationship between two tests (validity), the achieved scores should reflect considerable similarities if the examinee was not simulating his or her responses. Currently, the SB:FE (Thorndike, Hagen, and Sattler, 1986a) and the WAIS-R (Wechsler, 1981) are two of the most valid, reliable, and frequently used intelligence tests in the United States (Anastasi, 1988; Sattler, 1988). These tests are also frequently used to measure intelligence in criminal cases where the defendant's competency may be an issue (Grisso, 1986a; 1986b).

Statement of the Problem

There is concern that people with mental retardation experience injustices when they are involved in the criminal justice system (Thornburgh, 1992). People with mental retardation constitute a relatively small segment of the population of the United States, and estimates vary somewhat as to the actual incidence and prevalence rates, depending upon one's definitional perspective (Patton, Beirne-Smith, & Payne, 1990). If IQ were the only criterion, approximately 2.3 percent of the population would be

considered mentally retarded. The actual rate, however, has been estimated as somewhere between 1 percent (Tarjan, Wright, Eyman, & Keeran, 1973) and 3 percent (President's Committee on Mental Retardation, 1970).

Testing incarcerated and adjudicated individuals to determine functioning levels of intelligence has been recommended as a prudent step for defense attorneys (Burr, 1992; Ogloff, 1990). The number of people with mental retardation who are incarcerated has been estimated to be disproportionate to the proportion of such individuals in the general population (Ellis & Luckasson, 1985; Everington, 1987; Noble & Conley, 1992).

Noble and Conley (1992) reviewed a number of studies on the incidence and prevalence of mental retardation in U.S. penal institutions. For example, Brown and Courtless (1971), in a comprehensive review of over 90,000 inmates nationwide, reported approximately 9.5 percent of reported IQ scores were beneath the 70 cutoff. In addition, Denkowski & Denkowski (1985) reported wide differences of prevalence of mental retardation within penal institutions (1.5 percent to 19 percent), based upon type of test used for determination. However, a general rate falling between 2 percent and 6 percent was noted. Noble and Conley (1992) place the final estimate between 2 and 10 percent. The array of conflicting estimates may be attributed to many factors, including differences in testing procedures, test validity and reliability, population differences from state to state, and differences in each state's incarceration procedures. One clear factor emerges, however: The percentage of people with mental retardation in prison is still largely undetermined (Noble & Conley, 1992).

Few attorneys have had experience with people with mild mental retardation, it is questionable whether they recognize deficits in intellectual functioning. This is especially true in light of evidence that people with mild mental retardation will go to great lengths to conceal their handicap (Edgerton, 1967). As Resnick (1984) states, "Prisoners find the stigma of mental illness far worse than that of criminality" (p. 23). Given Edgerton's work, this fact is probably true for prisoners with mental retardation as well.

The possibility that incarcerated individuals might accurately simulate mental retardation on two tests in an effort to reduce their culpability before the court has not been investigated. If two examiners disagree on the actual diagnosis, the question arises as to which should be believed. Conflicts involving differences in professional interpretation within the criminal justice system warrant more proof of malingering than one person's professional opinion. Also, while trained psychologists may technically be capable of diagnosing mental retardation, and their views on possible evidence of malingering should be considered, the Court, even as guided by the expert testimony, may be essentially unqualified to make such a diagnosis. Other factors must be explored, particularly the mental retardation expertise of the witnesses (see generally, Conley, Luckasson, & Bouthilet, 1992).

Purpose and Rationale of the Study

The purpose of this study was to examine factors associated with the possibility of falsifying responses on corresponding subtests of both the WAIS-R (Wechsler, 1981) and the SB:FE (Thorndike, Hagen, & Sattler, 1986a), and obtaining standard scores that were not statistically different. While the resulting data do not reveal IQ scores, the profile obtained from the subtests administered constitute an important step in concurrent validation of the subtests utilized, as well as an expansion of the literature on malingering and deception.

To clarify exactly why it is difficult to obtain similar scores on these two similar subtests, it is important to understand certain factors of each test. The scoring procedures are different for each test. On the Comprehension subtest of the WAIS-R, examinees may obtain full, partial, or no credit for their response to only 16 questions. However, on the SB:FE, examinees are scored pass/fail on 42 questions, including various pictorial identification items not found in the WAIS-R. Resulting raw scores for each are translated to standard scores; however, these standard scores are calibrated on different scales. Scaled scores for subtests on the WAIS-R have a mean of 10 and a standard deviation of 3; scaled scores for subtests on the SB:FE have a mean of 50 and a standard deviation of 8 (Sattler, 1988).

If examinees deliberately and consistently attempt to falsify their responses on test items of different tests, there would necessarily exist some variation between the obtained results. For example, item scoring procedures vary between each test, making it unlikely that one could obtain similar scores on both. In addition, given inherent differences between the WAIS-R and the SB:FE, it is extremely unlikely that examinees could consistently falsify their responses so as to appear to function within the same range of mental retardation on both tests, and have enough similarity in the obtained subtest scores and IQs to avoid statistical differences.

Conflicting expert testimony and subjective opinion may be the sole determinants to the dispute over whether or not a defendant has simulated in responding to the tests. One conjecture is that

psychologists may presume that the research in other fields is applicable to persons with mental retardation. While research exists concerning simulating on personality tests in effort to feign mental illness (Benton, 1945; Hawk & Cornell, 1988; Ogloff, 1990; Rogers, 1988), and simulating on functional tests to determine specific sensory skill levels (Benton & Spreen, 1961; Pankratz, 1979; Pankratz, Fausti, & Peed, 1975), there is a paucity of supportive research on simulating on intelligence tests to present one's self as mentally retarded (Goebel, 1983; Heaton, Smith, Lehman, & Vogt, 1978; Rogers, 1984; Schretlen, 1986; Schretlen & Arkowitz, 1990; Spreen & Benton, 1962).

This study adds to the literature in this area and may serve as an important step to establish a basis for suggesting that mental retardation cannot be consistently simulated on tests of intelligence. This issue is important since case law determining the level of culpability and legislation regarding capital punishment for persons with mental retardation is still evolving (Burr, 1992).

Studies validating intelligence tests may consider various factors that often appear to be completely extraneous. Paradoxically, none have considered response falsification as a possible method to prove test validity. By identifying statistically significant differences between corresponding subtest scores obtained during intentional malingering, this study may also enhance the validity data on the subtests utilized herein. In addition, the possibility of performing consistently on these tests when answering the items genuinely will contribute to the literature on each test's concurrent validity.

Previous research studies have considered a response latency factor in deliberate deception (Goldstein, 1923; Langfeld, 1921; Marston, 1920). Such research investigated subjects' ability to deceive the examiner successfully without requiring increased time to develop a deceptive response. In these studies, results varied widely, and conclusions were often conflicting. In the current study, subjects' response times were investigated to assess response time differences between the control and experimental conditions.

Finally, this study surveyed subjects about insights they held as to the ease or difficulty they experienced in the different experimental conditions. In a debriefing session, subjects were requested to answer several questions designed to elicit their opinions, feelings, thoughts, and ideas about what they had attempted to do in order to determine the amount of agreement between subjects.

REVIEW OF THE LITERATURE

The research that forms the basis of this study may be categorized into two major areas. First, validity research of the test instruments is considered. Second, research on falsification of test responses is reviewed.

As previously stated, there are several forms of test validity. The primary form of validity under consideration in this study is concurrent validity, a form of criterion-related validity. Research on the WAIS-R (Wechsler, 1981) and the SB:FE (Thorndike, Hagen, and Sattler, 1986b) demonstrates that the degree of concurrent validity between these two tests is sufficiently adequate to indicate that these instruments measure similar attributes.

The vast majority of research in the area of simulation has concentrated on feigned mental illness on personality tests, not feigned mental retardation on intelligence tests (Lees-Haley, 1986; Rogers, 1988). This review investigated the need for research into simulating on tests of intelligence for purposes of identifying individuals who truly have mental retardation. Furthermore, specific implications relating to individual performance, test validity, and possible malingering within the criminal justice system were explored.

Test Validity

The concept of test validity is extensive and complex (Wainer & Braun, 1988). It refers to "...the extent to which [a test] measures what it purports to measure" (Wainer & Braun, 1988, p. 19). The statistical concept of a 'test's validity' may not necessarily be the real quest of a researcher or examiner; rather it is the validity of the inferences and information to be derived from such instruments. Various forms of validity determine the accuracy, and therefore the value, of the information obtained from testing instruments (see Table 1). The primary form of validity investigated in this study was concurrent validity, one of two kinds of criterion-related validity, the second being predictive validity. Concurrent validity refers to "whether test scores are related to some currently available criterion measure" (Sattler, 1988, p. 30).

Table 1. Forms and Definitions of Validity (adapted from Luckasson & Keyes, in progress)

BEST COPY AVAILABLE

Form	Definition
Face	Whether the test measures the actual domain which it claims to measure.
Content	Whether the items on a test adequately represent the domain which the test claims to measure.
Criterion-Related	How well a test relates to an outside criterion, such as another test which measures the same or a very similar concept.
Concurrent	A form of criterion-related validity which compares an instrument to an outside criterion, such as a teaching procedure that renders similar results. High accuracy in identifying a similar construct means the test probably has good concurrent validity with the procedure.
Predictive	Also a form of criterion-related validity, determines how accurately the test can predict a future outcome. An example is the SAT test, which claims to predict student success in college.
Construct	Largely hypothetical, the extent to which a test actually measures a psychological construct or other theoretical concept.

Cronbach (1927) characterizes the evaluation of validity not as mere research, but as an on-going argument, with inherent pros and cons. Again, the instrument itself is not the primary target. The information obtained must be viewed through the realistic understanding that, whatever the instrument, there are necessarily weaknesses that exist in data gathered, i.e., error. Validation, according to Cronbach, "is never finished" (in Wainer & Braun, 1988, p. 5).

Validity data are traditionally expressed as a decimal coefficient, where perfect validity and no factor of error is expressed as 1.00, and no valid relationship is expressed as zero. Validity may also be expressed in an expectation chart, where the examinee's score on one test is used to determine the chance of receiving a corresponding score on another criterion, such as another test (Anastasi, 1988). This is a form of predictive validity, but it relates to concurrent validity because the scores achieved on the first measure may be compared with the scores of the second measure once it has been administered (Wainer & Braun, 1988). This concept is important, since expectancy charts can be developed based upon the results obtained on both test administrations.

The validity of a testing instrument is crucial to the use of the information it provides. For instance, if a test manual claims that the test measures ability in math, then the test would be expected to examine various mathematical concepts, numerical procedures, and the like. However, if the test itself offers only word problems that do not use Arabic numerals, then the validity of the test is affected. This is because the test must be read, written concepts must be converted to mathematical procedures, and calculations must be completed. Therefore, this test is measuring more than the examinee's ability in math, thus limiting the test's validity for mathematical aptitude.

Research on individual test instrument validity is extremely important, since a claim by an examiner that a score measures some attribute, unless supported by validity data, is mere opinion. Obtained validity standards have traditionally been set at or between the .05 and .01 levels of statistical significance. Validity coefficients lower than .80 are often difficult to interpret, since they may indicate a weakness in a specific factor but still be relatively valid to the concept in general. This is especially true for construct validity, because a hypothetical construct, such as intelligence, is not yet conclusively defined and therefore is essentially incapable of being proven valid (Anastasi, 1988).

Test administration factors also affect a test's validity in the performance of individual examinees. Sattler (1988) states that:

> These include test taking skills, anxiety, motivation, speed, understanding of test instructions, degree of item or format novelty, examiner-examinee rapport, physical handicaps, degree of bilingualism, deficiencies in educational opportunities, unfamiliarity with the test material, and deviation in other ways from the norm of the standardization group." (Sattler, 1988, p. 31)

When considering the overall validity of any testing procedure, numerous factors may potentially affect an examinee's performance either positively or negatively (see Table 2). Various environmental stimuli, emotional and technical factors associated with the examiner, the examinee, and the circumstances of test administration may radically alter the validity of test performance. For example, if the examinee had no breakfast, he or she might concentrate on hunger and be unable to concentrate fully on the tasks presented, resulting in a depressed performance. Likewise, if the examination room is not well lighted, the examinee's performance might not be optimum, and the examiner's ability to observe the performance might be equally hindered.

Research investigating the potential of individuals to perform above their actual ability levels has been questioned (Sattler, 1988). However, research designed to determine individual ability to perform below expectancy has been viewed as an aspect of test validity and examiner experience (Caplan, Lubin, & Collins, 1982; Ironson & Davis, 1979; Loo & Wudel, 1979). This research has focused particularly upon personality tests, especially those which are designed to identify various personality traits considered to be deviant, such as the Minnesota Multipluralistic Personality Inventory (MMPI) (Dahlstrom, Welsh, & Dahlstrom, 1975).

Table 2. Factors Affecting Test Performance (adapted from Sattler, 1988).

POSITIVE	NEGATIVE
Good Health	Poor Health
Strong Educational Background	Weak Educational Background
Wide Range of Experiences	Limited Range of Experiences
Travel Experiences	No Opportunity to Travel
Positive Home Environment	Poor Home Environment

Table 2. Factors Affecting Test Performance - Continued, (adapted from Sattler, 1988).

POSITIVE	NEGATIVE
Good Use of English	English as a Second Language
Functioning Sensori-Motor	Sensori-Motor Deficits
Good Testing Environment	Poor Testing Environment
Comfort & Rapport w/ Examiner	Nervous & Unfriendly w/ Examiner

Validity data acquired from tests must also be viewed in the context of several factors relating to the nature of the group from which the data are obtained. Standardized intelligence tests are generally norm-referenced, meaning that they have been previously administered to a large sample of the population, and the resulting scores compared, computed, and transformed to fit obtained norms. The validity data of the two subtests administered in this study are assumed to be representative of the population at large, as initial normative and standardization information for each was based upon demographically correct cross-sections of the population, as determined by the United States Census Reports in 1972 (Wechsler, 1981) and 1980 (Thorndike, Hagen, & Sattler, 1986a; 1986b).

Despite consistent effort to provide accurate and valid scores during the standardization procedure, as is true with all data collected in scientific research, there exists a margin of error that must be considered. In validity data, this margin of error is referred to as the standard error of estimate (SE_{est}), also expressed as a coefficient. This coefficient expresses the estimated error, statistically determined through the obtained data, which is expected to occur between the different measures being utilized. Anastasi (1988) states, "...the error of measurement indicates the margin of error to be expected in an individual's score as a result of the unreliability of the test" (p. 168). This statistic is important to understanding that the performance of the examinee may not present an exact illustration of his or her abilities due to the structure and content of the test itself, not just the examinee's capacity. In addition, the smaller the standard error of estimate, the more valid and reliable the test (Anastasi, 1988; Sattler, 1988).

The validity of two tests may be estimated by the consistency with which an examinee responds to the two tests. Taking this to represent a factor of the test's construct validity, it is described by Grisso

(1986b):

> One of the most important uses of construct validity research will be to provide an empirical base for dealing with questions of response
> distortion: that [sic] is, malingering or dissimulation. Several strategies for detecting response distortion have been used by psychological examiners. Some of these are built into tests and produce special scores... Others rely on subjective examination of interview and test responses for unusual or incongruent patterns of response... detection of logical discrepancies between responding on two correlated instruments... or observation of discrepancies between behavior in assessment settings and other settings. Whatever the method, the reasoning underlying detection of response distortion relies on the known or assumed correlation between two behaviors. When an examinee manifests a pattern of behaviors that consistently violates these expected correlations between test responses and other behaviors (or between responses within a test), then the usual meanings of the test scores cannot be assumed. (p. 49-50)

Validity of the *WAIS-R* and *SB:FE*

Due to the age of the WAIS-R (Wechsler, 1981), an abundance of validity research is available. The amount of research on the SB:FE (Thorndike, Hagen, & Sattler, 1986a) is rapidly expanding, despite its relative newness. Both tests have been shown to hold impressive validity standards (see Tables 3 and 4) (Sattler, 1988).

Validity data available on the WAIS-R (Wechsler, 1981) and the SB:FE (Thorndike, Hagen, & Sattler, 1986a; 1986b) support the use of these tests to determine cognitive abilities. Tables 3 and 4 report validity information on both the WAIS-R and the SB:FE, respectively, and indicate that the correlation of scores between the global scores (Full Scale IQ and Composite Score) on these two tests is very high (.85 to .91).

Table 3. Concurrent Validity Research on the WAIS-R (adapted from Sattler, 1988)

Criterion Measure	WAIS-R Full Scale IQ
WAIS (1955)	.94
SB:FE (1986)	.85
Slosson Intelligence Test (1983)	.78
Revised Beta (1957)	.43
Woodcock Johnson Cognitive (1977)	.69
Years of Education	.54

Table 4. Concurrent Validity Research on the SB:FE (adapted from SB:FE Technical Manual, 1986b).

Criterion Measure	SB:FE Composite
Stanford-Binet, L-M Edition (1972)	.81
WAIS-R (1981)	.91
Wechsler Intelligence Scale for Children (WISC-R)	.83
Kaufman Assessment Battery for Children (K-ABC)	.89
Wide Range Achievement Test (WRAT-R)	.51-.58

The validity coefficients reported above vary somewhat, largely according to the type of test

instrument. When compared to other intelligence tests, the correlations range from good to adequate. It is important to note that among the lowest correlates for each test was years in school and the Wide-Range Achievement Test (Jastak & Wilkinson, 1984). This is not only appropriate, but desirable, since tests of cognitive capacity, or intelligence, should not correlate too closely with achievement because achievement tests are supposed to measure learned skills in order to determine educational success. High correlations with these factors would suggest that the WAIS-R and the SB:FE did not accurately measure innate cognitive abilities, but measured material that could be learned in school or in life experience (Sattler, 1988). In summary, both of these tests appear to measure similar attributes that are assumed to be essential to intelligence.

Among the most important research data available on both the SB:FE and the WAIS-R is information evaluating the subtests as measures of general intelligence factors. This research relates to a major theory of intelligence as a two-factor theory (Spearman, 1927); one which measures the "general" or "g" factor of intelligence, and one which measures a specific intelligence factor related to an individual skill. Factor analytic studies have determined the percentage proportion of the subtests' variance which measures "g." The subtests utilized in this study (Comprehension) are both considered to be reasonably good measures of "g" with median loadings ranging from .75 (SB:FE) to .78 (WAIS-R). The proportion of variance attributed to the "g" factor on both of these subtests was over 50%, suggesting that both subtests measure aspects that are highly related to general intellect (Sattler, 1988). This means that, when answering Comprehension questions, the examinee's general intellect is utilized significantly more than in other subtests that measure different characteristics of intelligence. The factor analytic findings have significant bearing on the generalizability of the results of this study.

Test Validity and Criminal Justice

The concept of validity can be confusing. Information obtained by test instruments, particularly the use of standard scores, must be considered in light of conditions which affect validity. Grisso (1986a), in referring to the use of such tests with incarcerated individuals, states:

> For most psychological tests, the norms with which the psychologist will compare the examinee's performance will not have been collected under psychological conditions similar to those in which the present assessment is being conducted. Very few tests have been normed on persons who have just spent their first two days in a crowded city jail, have just been issued prison clothes in exchange for their own, or are waiting judgment... The psychologist must take special care to be aware of the effects of such situations when observing the examinee's behavior during the assessment session, when administering various tests, and when using normative data to interpret them. (p. 123)

This statement identifies the importance of clinical judgment when using normative data in extraordinary situations. As in any clinical testing situation, various methods of obtaining psychological data must be utilized in effort to obtain an adequate profile of the examinee. Clinical judgment is required to piece together the body of data. Grisso (1986a) has suggested that one interpretation of conflicting information obtained between behavioral observations (either reported or directly observed) and test performance may be that "one's test data reflect motivated attempts by the examinee to 'look good' or to 'look bad'" (p. 123).

Malingering and Deception

Defining the Concept

Three approaches to malingering have been suggested: Separate from mental disorder, symptomatic of mental disorder, and nonexistent. The American Psychiatric Association distinguishes between malingering and mental disorders, and notes that the major feature is

> "the voluntary production and presentation of false or grossly exaggerated physical or psychological symptoms...(p)roduced in pursuit of a goal that is obviously recognizable with an understanding of the individual's circumstances, rather than of his or her individual psychology." (DSM III-R, 1987, p. 23)

Diamond (1956) and Menninger (1963), on the other hand, supported the concept of malingering as symptomatic of an underlying mental illness, since individuals who go to great extremes to convince others

that they are ill are clearly mentally abnormal. Other clinicians have suggested that, given the unconscious nature of thought, malingering cannot, essentially, exist (Benton, 1945). In their opinion, no thought is completely conscious, and thus the intent to pursue a goal is involuntary, that is, the malingerer cannot help but malinger. However, conscious or unconscious, according to medicolegal case law, malingering implies intent (Gorman, 1982; Swanson, 1984).

The clinician's ability to detect deception has been questioned, largely as a result of varying definitional factors (Faust & Guilmette, 1990; Faust, Hart, Guilmette, & Arkes, 1988; Kuvin, 1986; Markus, 1986; Rogers, 1990a). Topics such as the level of training required to diagnose malingering, the determination of motives, and the observational, testing, and anecdotal information necessary to diagnose malingering have all been questioned as a result of the seemingly vague nature of the DSM III-R definition (Bigler, 1990; Eisner, 1985; Faust & Guilmette, 1990).

It is probably safe to assume that as long as there have been tests, there have been people who tried to cheat on them in order to achieve false results (Flicker, 1956). Examinees who tried to present themselves as something other than they are have been researched in several areas, especially in psychology and criminal justice (Grisso, 1986a). While there is a lack of research in the general area of malingering (Ziskin, 1984), certain forms of malingering have received more attention in the literature (Anderson, Trethowan, & Kenna, 1956; Ogloff, 1990; Rogers, 1988).

Szasz (1956), in a review of research related to the diagnosis of malingering, strongly disputed malingering as a diagnostic category. Noting difficulties in medicolegal, psychopathological, criminological, and sociopsychological areas, he described diagnosing 'malingering' as one would diagnose a legitimate mental illness such as schizophrenia, as "a grave error" (p. 432).

Szasz's essay has very important theoretical implications. Szasz noted that the very definition of diagnosis supported his argument. He states:

> Any given diagnosis has, at the very least, the following... functions...: 1. It represents the physician's (psychiatrist's) concept of what is "wrong" with the patient (e.g., fractured leg...). 2. It serves as a method of communication between him and other physicians. That is, each 'diagnosis' calls for appropriate 'treatment'. (p. 433)

The argument centers upon the moral condemnation of malingering as being, essentially, lying, and thus differs from legitimate conditions which do not necessarily carry a conscious factor (i.e., schizophrenia). To "diagnose" someone of lying becomes more of a judgment call than a medical determination, and should not carry with it any possibility of subsequent rewards, i.e., compensation or relief from duty.

Suspending any notions of psychiatric diagnosis, Szasz views the malingerer as someone who considers life, or certain aspects, a game, and as such, views malingering as "cheating" (p. 434). The determination of actual malingering, therefore, is the essence of the problem, since

> ...not long ago all sorts of behavior now regarded as 'psychiatric illness' was thought of as 'malingering.' Even today, those unsympathetic to the psychiatric mode of thought concerning human behavior tend to think (and speak) of deviant behavior as 'malingering.' (p. 435)

One may question the logic of Szasz's argument, however, if malingering is still considered by the psychiatric community at large as a legitimate psychiatric diagnostic entity, almost 35 years after it was originally set forth. Perhaps the most appropriate consideration would be to view the argument as it relates specifically to those potential servicemen attempting to evade active duty in times of war, i.e., cowardice. Given this aspect, it may also be related to those defendants attempting to avoid potential execution by faking mental illness, or, in certain situations, perhaps mental retardation.

Rogers (1990a) supports the use of convergent validation (use of more than one test) in determining malingering. He suggests that the malingerer perceives his or her situation as adversarial, and that malingering is chosen as a logical and utilitarian adaptation. Rogers questions the value of the DSM III-R approach, maintaining that it has an "unduly moralistic" (p. 182) focus. In its stead, he proposes models based upon empirical evidence, adaptational qualities, and decision theory.

The feigning of mental retardation has received very little research attention (Flicker, 1956; Ogloff, 1990), and the use of intelligence tests as a measure of control for diagnostic validity has not been investigated. The feigning of mental illness has received more investigation (Ogloff, 1990). Feigning various forms of personal injury for financial compensation has been well researched, as has the simulation of brain injury, although not with the intent of proving mental retardation (Bruhn & Reed, 1975; Goebel, 1983).

Due to the dearth of research in the malingering of mental retardation, much of this review will investigate malingering of mental illness and attempt to draw parallels to the malingering of mental retardation. However, it is necessary to distinguish between the legal definitions of these two conditions. A person with mental illness is referred to as one "who suffers a substantial disorder of thought, mood, perception, orientation, or memory which grossly impairs judgment, behavior, or the capacity to recognize reality or the ability to meet the demands of life" (American Bar Association, 1989, p. 467). A person with mental retardation is referred to as one with "...significantly subaverage general intellectual functioning existing concurrently with deficits in adaptive behavior, and manifested during the developmental period" (Grossman, 1983, p. 1). Research on falsification of test responses is divided into three sections: Characteristics, mental retardation, and mental illness.

Characteristics of Malingering

According to the American Psychiatric Association (DSM III-R, 1987), falsification of responses may occur for several reasons (see Table 5). However, the most likely reason would appear to be based upon the examinee's perception that there might be some form of personal gain by performing poorly on the test (Lezak, 1983).

One characteristic essential to malingering is the recognized potential for secondary gain. The individual who consciously malingers must be aware, at least to some extent, of the possibility of a payoff. If an examiner is suspicious that the examinee may be simulating, he or she should look for the existence of positive reinforcers for the deception (McGarry, 1986).

Table 5. Factors Producing Malingered Responses (adapted from DSM III- R, 1987)

1. Job screening (avoiding work)
2. Military justice situations (avoiding conscription or duty)
3. Criminal justice situations (evading prosecution)

Table 5. Factors Producing Malingered Responses - Continued(adapted from DSM III- R, 1987)

4. Psychological disability payoff (financial compensation)
5. Physical disability payoff (financial compensation)

Several characteristics and predictors are useful in identifying malingering. It is apparent that in a large number of cases where malingering is suspected, the purpose is likely to be for financial gain. Despite this, malingering is still considered to be a personality disorder characterized by a lack of honesty. For example, the Clinician's Handbook (Meyer, 1983) states:

> [Malingering] occurs more commonly in the early to middle adult years, is more common in males than in females, and often follows an actual injury or illness. Problematic employment history, lower socioeconomic status, or an associated antisocial personality disorder are also common predictors of this *disorder*. (p. 360 - italics added)

In this respect, the determination of malingering must be made in light of the fact that the act of malingering is, in and of itself, maladaptive, or possibly indicative of other personality or organic problems (Flicker, 1956; McGarry, 1986).

Sierles (1984), using an anonymous questionnaire, concluded that sociopaths, drug abusers, and alcoholics were more prone to malingering than other individuals. No correlates were found, however, between malingering and somatic disorders, or disorders that are actually physical. This finding suggests that somatic disorders should not be equated malingering, despite assertions to the contrary (Benton, 1945).

Swanson (1984) contends that malingering is "conscious, voluntary, and goal-directed behavior" (p. 287) that is always false and fraudulent. Citing DSM III, he notes that a "clearly definable goal" (p. 287) must be present that differentiates the malingerer from factitiously ill subjects, or those subjects who truly believe that they are ill, although they are not. Additionally, any combination of antisocial personality disorder, medicolegal aspects, discrepancies between objective findings and the patient's claims, and lack of cooperation in diagnostic or therapeutic recommendations should increase the physician's

BEST COPY AVAILABLE

suspicion of malingering (DSM III, 1983; Swanson, 1984).

Diagnosis of malingering appears to be a difficult task for even the most experienced psychiatrists and psychologists (Bigler, 1990). Early in the research, Williams (1931; reported in Flicker, 1956) indicated that even experienced clinicians were less likely to diagnose malingering, pointing out that psychiatrists have "an unfortunate tendency" (p. 24) for finding everyone to be abnormal. Despite more recent work on mental health standards in criminal justice, this issue is still a difficult problem within the legal system.

Pankratz and Erickson (1990) considered two views of malingering. First, they suggest that clinicians may want to avoid labeling patients they apparently could not help, as it would appear to serve no purpose, and may lend credibility to a possible deception. Second, they suggest that avoiding the diagnostic process may result in disadvantages since clinicians may appear to be neglecting patients who are in need. The technology available for the determination of deception, the patient's legal situation, and the patient's intentions are regarded as highly important to the accurate determination of malingering. Hartings (1989) suggests that a clinician's familiarity with various aspects of organic and psychiatric symptoms that the suspected malingerer may feign, as well as the clinician's skilled use of techniques to probe the patient's suggestibility may increase accuracy of detection.

As stated, Rogers (1984; 1990b) offers an empirical model of characteristics to identify dissimulation based upon response styles of the subject (see Table 6), and a model of adaptive responses to adverse circumstances based upon decision theory. The complexity of diagnosing malingering requires that symptoms associated be chronicled, and, by synthesizing the resulting data, accurate diagnosis will be enhanced. Three forms of data are outlined: Case studies, psychological testing, and social psychology research. In addition, the determination of malingering must consider two factors, psychopathology or criminal background. Combining clinical data with corroborative evidence will reduce the chance of erroneous diagnosis.

Table 6. Empirical Model of Dissimulation (adapted from Rogers, 1984).

Indicators	Data form
1. Extreme severity of symptoms	Case study
2. Consistency of self-report	Psychological testing
3. Likely rare symptoms	Psychological testing
4. Inconsistent symptoms sequence	Case study
5. More obvious symptoms	Psychological testing
6. Sudden onset of symptoms	Case study
7. Likely admission of common foibles	Psychological testing
8. Unlikely idealistic self-attributes	Psychological testing
9. Patterned responses	Psychological testing
10. Observation/symptom conflict	Case study
11. Possible high failure rate	Psychological testing
12. Nearly correct responses	Psychological testing
13 Willingness to discuss symptoms	Case study
14. Stereotypical neurosis possible	Psychological testing

Three other factors outlined in Rogers' model, but not specifically linked to malingering, include vague responses, fidgetiness, and latency of response. The latter factor, response latency, appears consistently throughout the early research on malingering. Latency refers to the amount of time that exists between the cessation of the stimulus and the onset of the response (Alberto & Troutman, 1986). This factor is a critical aspect of the present study.

Occasionally, clues of possible malingering can occur in very unlikely ways. Rosenham (1972) and Ritson and Forest (1970) identified separate instances in which hospitalized psychiatric patients apparently have identified fakers, even to the extent of having told malingerers to "stop play-acting" (Resnick, 1984, p.30). To delineate specific methods of identifying such malingerers, Resnick (1984) suggested sixteen clues to malingered psychosis (see Table 7).

Motivation factors associated with malingering studies have included monetary payment (Bernard, 1990). Performance on four neuropsychological memory instruments was tested using college students assigned to three groups; control, malingering with a financial motive, and malingering without a financial motive. Results indicated that the financial motive did not result in difference in performance. However, both malingering groups did perform significantly poorer on the tests administered. Using discriminative functions of the combined results as predictors, nearly 75% of malingerers were identified. This may, however, still be considered as a high rate of error, since 25% of the malingerers were not identified. Bernard suggested that neuropsychological memory tests may be vulnerable to malingering.

Table 7. Clues to malingered psychosis (adapted from Resnick, 1984).

1. Overacting; extremely bizarre behavior; physical representations
2. Eager to call attention to the illness; schizophrenics don't
3. Difficulty in imitating form of illness, but not content; too exact
4. Symptoms that don't fit a known diagnostic entity
5. Sudden onset of delusion; schizophrenic delusions take weeks
6. Behavior doesn't conform to delusions; atypical of schizophrenia
7. Makes up delusions to fit the facts of a crime
8. Laughs or embarrassed when exposed to discrepancies in story
9. Present themselves as blameless due to their 'illness'
10. Slow in responding; takes time to make-up a response
11. More than one person involved in criminal action
12. Ulterior motive for behavior; nonpsychotic alternative motive
13. Rarely show perseveration; most psychotics do perseverate
14. More likely to describe auditory hallucinations; "Go kill"
15. Unlikely to show residual schizophrenia; affect, peculiar thought
16. Real schizophrenics may also malinger auditory hallucination

After specific incidents of trauma, the malingerer also may appear to have various conflicts of behavior and identifiable forms of psychological distress (see Table 8). Such observations should alert the trained clinician that there is a possibility of deception. Although these inconsistencies may occur in people with true mental illness, combinations or unusual persistence in any of these behaviors may alert the clinician to the possibility of malingering (Resnick, 1984).

Table 8. Inconsistencies and Characteristics of Malingerers (adapted from Davidson, 1965; Kilgore, 1982; and Resnick, 1984).

1. Unable to work, but enjoys recreation, games, television, theatre
2. History of drifting and spotty employment
3. Evasive and unwilling to discuss work, finances, expectations
4. Sullen, ill-at-ease, suspicious, uncooperative, or resentful
5. Avoids examination, unless required for financial benefit
6. Declines to cooperate in diagnostic or therapeutic procedures
7. History of incapacitating injury or extensive absences from work
8. Greedy, dishonest, unpleasant, or demanding personality
9. Marginal member of society for many years
10. Depicts self and prior functioning in exclusively glowing terms
11. Pursues a claim tenaciously, despite depression or incapacitation
12. Refuses employment despite ability

Assessment of Feigning

Gorman (1984) contends that malingering is "a legally wrongful act, which can best be diagnosed by a physician's examination" (p. 67). He outlines a two step process: First, proving that the symptom is not due to a diagnosed problem; and, second, proving that the subject's current circumstances can account for the symptom.

Patterns of performance can be determined by the use of most tests of personality and intelligence

(Grisso, 1986b; Lezak, 1983). Such patterns may make it possible to determine the existence of deception in testing. Several tests of personality offer scales designed to identify examinees who are lying in their responses. The Minnesota Multiphasic Personality Inventory (MMPI) (Dahlstrom, Welsh, & Dahlstrom, 1975) offers several scales that are alleged to distinguish those who are giving honest responses from those who are lying. This instrument has received considerable research, particularly the so-called 'lie scales,' the L, F, K, F-K, and O-S validity indexes (Grossman & Wasyliw, 1988; Hawk & Cornell, 1989; Parwatikar, Holcomb, & Menninger, 1985; Roman, Tuley, & Villanueva, 1990; Schretlen, 1988; Walters, 1988; Wasyliw, Grossman, Haywood, & Cavanaugh, 1988).

Two of these scales are specifically designed to identify individuals who are "faking good" or "faking bad," in order to present the appearance of deviation. The alleged success of these scales in identifying deception has been both hailed (Hawk & Cornell, 1989) and decried (Roman, Tuley, & Villanueva, 1990).

Hawk and Cornell (1989) used the MMPI in an attempt to assess 18 malingering inmates, 17 psychotic defendants, and 36 control subjects on various personality factors. They reported that about 50% of the malingering and the psychotic defendants they attempted to assess were untestable or produced incomplete protocols, primarily because of noncooperation. Cooperative subjects' profiles were compared and revealed significant differences between the malingering subjects and other groups on the F-K index analysis.

Drob and Berger (1987) proposed three criteria for establishing the existence of malingered mental illness: Observed classic signs and symptoms of malingering; determination of a motive; and, ruling out genuine psychopathology that would cause voluntary symptoms. Applying these criteria to four cases, they found one malingerer, one factitious disorder, one combination of these two, and one uncooperative subject.

Rogers (1990b) states that:

> The assessment of malingering and deception remains a paramount issue in the practice of forensic psychology and psychiatry. Clinicians are challenged by both civil and criminal cases, where the penalties for self disclosure and incentives for dissimulation exert strong forces on the evaluative process. In the face of such forces, it is surprising how many patients are honest and self disclosing, even to their own detriment. (Rogers, 1990b, p. 1)

The mention of incentives in this statement has significance to the research of variables that effect malingering. For example, Heaton, et. al., (1978) used financial reward for better performance in their research; however, they did not determine if this incentive had any effect on subject performance. The effect of such factors should be investigated in relation to the level of successful malingering.

Given the findings of prior research, it is clear that more than one clinical, observational, or technical method of testing is necessary prior to the diagnosis of mental illness or malingering. Once this is done, the validity of the findings can be determined by synthesizing the data in such a way as to prove or disprove the existence of malingering and/or illness.

Norris (1943) observed that the determination of whether or not a plaintiff or defendant is malingering is often left to others who are not clinicians:

> It might well be considered that a medical man, as such, has no special qualifications to decide whether his patient is guilty of fraud. In any case, if he is asked in court whether the claimant is a malingerer, it would be quite proper to reply that this appears to be a question for the court. (reported in Flicker, 1956, p. 29)

Early Research - Deception and Time Latency

The advent of psychoanalytic theory and associated experimental techniques introduced a number of combined methods for detecting deception, particularly the "word association" technique (Jung, 1910). In an effort to determine methods of identifying those who would deceive on such tasks, Yerkes and Berry (1909) noted that time factors could be measured, suggesting that reaction times tended to be longer for deceptive responses.

Disputing such theories, however, Marston (1920) suggested that reaction time differences, rather than being a global factor, could result from different personality types. Positive and negative typing was used to indicate whether or not subjects investigated were inclined to succeed or fail in the commission of the deception. Those subjects who were "positive" types showed increased reaction time for deception;

"negative" types showed no increased reaction times when deceiving.

Goldstein (1923), in an elaborate replication of certain aspects of Marston's study, found that negative type subjects were not only less inclined to decreased reaction times, but that they apparently did not experience difficulty in making the decision to consciously deceive (referred to in the study as "disobey") in their responses. Positive type subjects exhibited opposing behaviors, presumably cognitive in nature, illustrated in the statement of one subject who, when asked to describe his reactions, stated:

> When I know that I am going to obey, I feel relieved and relaxed and I go with much more speed. In disobeying, it is harder to get adjusted and to concentrate and I am much slower in my responses. I hesitated and stopped at intervals when disobeying and was more conscious of the fact and was more excited. It didn't go smoothly. There was more of a strain at this time. (Goldstein, 1923, p.570)

Conclusions reached in this study initially supported a 'personality type' theory like that suggested by Marston (1920). However, Goldstein's initial findings suggested that negative type subjects were not actually conscious of the deception in their disobedience. She states that, "The so-called negative type is not a type of response to deception, but a type of non-deceptive response to this particular experimental situation" (p. 573). In short, these individuals did not feel they were being deceptive because they had been previously instructed to disobey instructions, and thus did not experience discomfort when not following subsequently supplied instructions. Positive type subjects, however, did report increases in strain, conflict, and emotional disturbance when attempting to deceive by disobeying supplied instructions. When instructions were worded or altered to suggest clear deception, however, almost all subjects took on positive type characteristics in their responses, and reaction times were universally increased.

In a second experiment, Goldstein altered the instructions and the experimental situation to clarify the subject's consciousness of the deception, and earlier differences between the positive and negative groups diminished. She stated that negative and positive

> ...types of response were, therefore, not two types of response to deception, but a deceptive and a non-deceptive type of response to the particular experimental situation.... In both experiments *consciousness of deception was accompanied by lengthened reaction times*, as compared with reaction times when [subjects] were not conscious of deception. (Goldstein, 1923, p. 580 - italics added)

The crucial concept arising from Goldstein's study is the determination that individual perceptions of deception vary across settings, and instructions. When subjects are conscious of deceiving others, instructions notwithstanding, reaction times tend to increase. However, when instructions clearly direct the subject to deceive, individuals respond according to their perception of the instructions; having been instructed to deceive, they may or may not be conscious of the *intent* to deceive, depending on the individual. This relates to the suggestibility of the individual subject, and has clear implications to the possibility of successfully malingering on tests.

Physical Manifestations and Time Latency

Langfeld (1921) combined various word associations and systolic blood pressure measures (a forerunner to the modern "Lie Detector Test") to determine the level of honesty at which subjects were responding to questions. Like Marston, he began with the concept that the subject's personality type may figure significantly in the ability of the researcher to determine the existence of deception. The assumption was that, if the subject was a "nervous" individual, outward signs of guilt (flushed cheeks, restlessness) would be observed, despite eventual determination of innocence. Conversely, a "stolid, self-possessed" subject may show no such symptoms, yet still be guilty. Stimulus words, half of which were considered "crucial" words that related to the "crime," were presented to both subjects, and associative responses were analyzed to suggest the subject's guilt or innocence. After a 3-day delay, blood pressures were obtained at various intervals during interrogation (Langfeld, 1921).

Several conclusions were drawn from the results obtained in Langfeld's procedure. First, the guilty subject's average reaction time to crucial words was considerably longer than reaction to control, or "non-crucial" words. Second, Langfeld determined that the time factor was a better indicator of guilt than the qualitative consideration of the actual response proffered in the word association procedure. Third, and perhaps paradoxically, the innocent subject appeared more nervous than the guilty subject. Finally, the innocent subject also lied during the blood pressure procedure, yet there was no rise in the systolic measure.

This supports the notion that the guilty subject's rise in blood pressure when lying on specific questions relating to the crime is caused by the conscious suppression of the truth. Individuals who lie may not be conscious of their own deception, and, as such, would not display physiologic traits associated with lying. When the subject, guilty or innocent, is conscious of the intent to deceive, associated physiologic traits tend to occur. Thus, consciousness of the intent to deceive would appear to be an important key to the success or failure of malingering.

The consideration of time factor variation in responses to crucial and non-crucial situations is significant. Still, results of this study were based upon the responses of only two subjects (one test trial), thus any suggestion of response generalization is questionable at best. It is, however, interesting to note that combining these two procedures in this experiment may have heralded the eventual combining of specific cognitive and behavioral aspects, major aspects of which are now known as elements of cognitive behaviorism (Bandura, 1969).

Malingering may be used to escape undesirable consequences or to achieve a specific goal, and specific efforts to achieve such purposes are well documented. Either of these situations may be viewed by the subject as having potential for secondary gain. For example, feigning mental or physical illness to avoid military service has been well documented throughout history (Benton, 1945).

During World War II, malingering was not uncommon among servicemen attempting to avoid active duty (Flicker, 1956). A wide array of studies and reviews were developed, espousing many, often conflicting views (Campbell, 1941; Gill, 1941; Hulett, 1941; Hunt & Older, 1943). However, once again the literature virtually ignored malingered mental retardation and concentrated upon malingered mental illness.

Benton (1945) investigated the performances of suspected malingering patients on the Rorschach Ink-Blot Test. These patients complained of various subjective physical illness or conditions: However, it was suspected that they did so solely in order to avoid active military duty. Again, delayed time responses were considered to be indicative of possible malingering. In addition, failure to supply the most common responses, general perplexity, meager interpretation, or complete lack of interpretation of even the simpler plates suggested less than accurate or honest performances. Careful clinical observation of the patient's inconsistent behaviors could result in the clinician's disregarding the existence of certain disorders characterized by specific symptoms. However, certain hysterical or experientially-based disorders, such as traumatic stress disorders, not uncommon during times of war, required further investigation.

Additionally, certain testing procedures, such as projective personality tests, may present patients with opportunities to deceive the examiner. Benton noted that, when a battery of tests was administered to a specific patient in effort to determine his or her mental and physical functioning, results of most tests showed average functioning and reasonable adjustment. However, when confronted with the Rorschach, the patient appeared to "smell a rat... [and presented a profile with]...a degree of constriction and/or poverty of ideation of almost psychotic proportions... quite inconsistent with all the rest of the patient's behavior" (p. 94 - 95). Whether it was the type of testing instrument or the different context in which the test was presented, the patient's response behavior appeared to change when presented with this task. In his conclusions, Benton states that

> ...it should be emphasized that this discussion concerns patients suspected of simulating certain physical complaints and does not necessarily apply to the problem of the simulation of mental defect or disease... A systematic investigation which would include a comparison of the performances of suspected malingerers... should yield results of considerable practical value and theoretical interest. (p. 96)

Feigning Mental Retardation

The question of faking mental retardation is not new, yet it does not appear to have received wide investigation in recent history. Wilbur (1852) described a condition appearing to be mental retardation but, in fact, was not retardation, as "simulative idiocy" (p. 35). He describes these children as those

> ...whose development has been retarded from congenital or other causes of a physical nature; and where these causes have been removed by the recuperative effort of nature, but the subjects are left bound down by the strong force of improper habits, which can be overcome only by the judicious labors of a suitable instruction. In these cases the result can be predicted with the utmost certainty. It will be the complete preparation for all the ordinary duties and enjoyments of humanity. (Wilbur, 1852, p. 35)

This was among the earliest references to what eventually became known as "pseudoretardation," an appearance, not actuality, of mental retardation.

While this may not be comparable to feigning, it is noteworthy, since behaviors indicated mental retardation where none existed, and suggests that sufficient training could significantly diminish the deleterious effects of these behaviors. However, this too, may be misleading; the question arises whether one may be 'trained' to act 'normally' when one is only 'simulating' abnormality in the first place.

Flicker (1956) identifies "oligophrenia" (arrested mental development) as a major reason for malingering. He states:

> If malingering is at all to be considered as an adaptive phenomenon, then certainly in [oligophrenia] it has its greatest excuse for being. [If] it is the individual of unsound mind who feigns psychosis, so ... it is the oligophrenic who feigns feeblemindedness... In many cases... individuals with definitely low intelligence will deliberately answer questions wrongly in order to give the appearance of even less intellect. (p. 26)

This quote, particularly the last sentence, is directly contrary to other research, most notably Edgerton's (1967) reports of a 'cloak of competence,' where people who have mental retardation attempted to "pass" for normal. The difference here appears to be in the assumption that, while the oligophrenic is not intellectually normal, he or she is not so handicapped by lower intelligence as not to see the advantage in "acting" more mentally retarded in a given situation. Conversely, the examinee who truly has mental retardation may so desire to be considered 'normal,' that he or she is unable to realize, in given situations, the potential advantages to being mentally disabled (Noble & Conley, 1992; Resnick, 1984). If there is a perceived potential for secondary gain, the malingerer may go to extraordinary lengths to deceive the examiner (Flicker, 1956).

The use of intelligence tests is the most common professional practice to identify persons with intellectual functioning within the range considered to be mental retardation (Noble & Conley, 1992). The most widely accepted definition was developed by the American Association on Mental Retardation (Grossman, 1983), and requires 'significantly subaverage general intellectual functioning' on an 'individually administered test of intelligence' (p. 3). This refers to an obtained IQ, or comparable global or composite intelligence score, that is two standard deviations below the mean. The determination of such a low score on an IQ test would likely result in further investigation, such as a second test administration, to reliably confirm the diagnosis. Feigning, in this context, would be extremely difficult, since each test is different, and the scoring procedures vary from test to test.

As stated, almost no research has investigated the possibility of faking mental retardation on tests of intelligence, yet statements of such possibilities do appear. Resnick (1984) opines:

> It is difficult for a person of normal intelligence to successfully fake mental retardation. Psychological testing is usually quite helpful. Intelligence testing is likely to show success on some difficult items and failure on some easy ones. School records, earlier psychological assessments, and military records should always be sought. A careful vocational history may belie serious intellectual deficit. Mild mental retardation does not prevent defendants from malingering auditory hallucinations or more severe retardation to avoid punishment. (p. 29)

The Standard Progressive Matrices test (Raven, 1960) has been used to investigate the possibility of 'faking bad' to give the appearance of cognitive deficit (Gudjonsson & Shackleton, 1986). Three groups, two with various forms of mental disorders and one malingering, were compared for performance similarity. Using a statistical method to determine the rate of decay across trials, the investigators were able to discriminate accurately between genuine and faked response protocols. However, it is important to note that, while the Standard Progressive Matrices test has been used to measure certain areas of intelligence, it is not, by itself, a sufficient indicator to determine one's actual functioning level (Anastasi, 1988).

In a study investigating individual subject ability on character set recall, Goldberg and Miller (1986) cite Lezak's (1983) hypothesis that only significantly deteriorated patients will remember less than three of five different character sets. Using a simple memory test, they tested 50 psychiatric patients and 16 patients with mild mental deficiency and determined that Lezak's criterion was accurate, that is, their subjects were capable of such a task. Thus, they support Lezak's theory, stating that individuals who deny remembering at least 9 of 15 items (the same proportion as Lezak's criterion using items, not sets) should be suspected as possibly malingering. This research suggests that memory tests, far less complex than intelligence tests, may be valuable in the diagnosing of malingering.

The plausibility of malingering on a test of intelligence in effort to prove mental retardation must be considered in another light. The definition of mental retardation outlined above (Grossman, 1983) also requires that deficits in adaptive behavior be present. Behavioral instruments used to assess adaptive behavior, often described as self-report measures or behavioral checklists, have often been seriously questioned in the literature (Zigler, Balla, & Hodapp, 1984), and the technical accuracy of such instruments has been criticized (Sattler, 1988). Still, adaptive behaviors, or adaptive skills relating to social maturity, remain fundamental to the definition and diagnosis of mental retardation (Barnett, 1986; Grossman, 1983; Luckasson, et al., in progress). The use of multiple respondents to gather accurate adaptive data is often necessary with this type of instrument. Such a procedure, a form of convergent validation, to determine adaptive functioning may be helpful in distinguishing between those who are malingering and those who are truly mentally retarded.

Another part of the current definition of mental retardation requires that the diagnosis be "manifested during the developmental period" (Grossman, 1983, p. 3). This means that in order for a diagnosis of mental retardation to be valid, the condition must appear prior to the person's 18th birthday. If a person should become brain damaged after their 18th birthday to such an extent that they might be considered retarded, the condition is then referred to as dementia (DSM III-R, 1987; Grossman, 1983).

The triple nature of the definition supplies an important requirement to any possible diagnosis of mental retardation. Also, records and historical data on the subject can offer valuable assistance to an accurate diagnosis (Miller & Germain, 1988). In any case, when the occasion for secondary gain is clear, malingering must be suspected, particularly when evaluating defendants (Resnick, 1984).

McGarry (1986) states: "Nowhere in clinical psychiatry are the skills and knowledge of the clinician more challenged, especially in legal settings, than in the diagnosis of malingering" (p. 83). Clinical psychiatry often involves the use of testing instruments to assist in the diagnosis of a patient's illness or mental problem. The use of intelligence tests and resulting IQ scores has been a controversial issue in the determining the existence of mental retardation (Noble & Conley, 1992) and of mental disorders (Ogloff, 1990) among incarcerated individuals. Conflicting opinions exist which suggest that people may or may not be able to successfully feign mental disorders or brain damage. Heaton, Smith, Lehman, & Vogt (1978) found no significant differences between IQs and test battery scores of subjects with actual disabilities and subjects who were asked to malinger, although the patterns of strengths and weaknesses varied widely between these groups.

Other research indicated that scatter analyses of subjects' performance on tests of intelligence to identify malingering has possibilities. Schretlen (1988) found evidence that test item discrimination may more accurately identify malingerers; that is, that malingerers will get certain items incorrect that actual disabled individuals will get correct, and vice-versa. Despite evidence that intelligence tests can be useful, Ogloff (1990) warns that

> (W)hile intelligence tests may eventually prove useful and accurate in identifying malingering and deception, it appears that extreme caution must be used when a clinician attempts to rely exclusively upon *an* intelligence test to identify malingerers. (p. 34 - italics added)

Schretlen and Arkowitz (1990) support the use of systematic and comparative testing to determine the existence of malingering. They state:

> It has also been shown that fakers can elude detection on a *single* test (Albert, Fox, & Kahn, 1980; Gough, 1947). Other studies suggest that a battery of tests on which response demands vary (e.g., structured versus unstructured) may detect faking more accurately than a single instrument (Bash & Alpert, 1980; Heaton, et al., 1978)... Finally, most investigators have applied intelligence tests to the detection of faked mental deficiency and personality tests to the detection of faked emotional disorders. However, several studies (Anderson, et al., 1956; Wachspress, et al., 1953) suggest that items from one domain may reveal faking in the other. If these findings represent a robust phenomenon, then intelligence test items may detect persons faking psychological disorders and personality test items may detect persons faking retardation. (p. 76; italics added)

The diagnosis of malingering may require the input of a psychiatric professional who is trained to accurately identify specific mental problems. However, the diagnosis of mental retardation requires more than just an IQ test administered by a psychiatrist who may or may not have had experience with this condition. It is possible that a highly trained professional psychologist or psychiatrist, having had no

experience with people who have mental retardation, could misinterpret retarded functioning on a test as malingering. Reputable diagnosis of any condition, injury, illness, or even deception requires validation (Ogloff, 1990).

Schretlen and Arkowitz (1990) tested two groups of prison inmates who were instructed to respond in ways that would identify them as retarded or insane. Subjects were administered a small battery of tests that included a Bender Gestalt, an MMPI, and a malingering scale developed for the study. Additionally, the researchers offered a monetary incentive for successful deception and included a criterion group of persons with mental retardation for comparison. The goal was to consider interactive effects in order to identify those who were faking. While they found that 80% of subjects instructed to fake mental retardation were exposed, they also determined that those instructed to act mentally retarded presented themselves as emotionally disturbed, and vice versa. While this study supports the use of multiple instruments to determine malingering, it did not utilize an individual test of intelligence.

Difficulties in Feigning Mental Illness

Due to the generally accepted standard in criminal justice which prohibits 'blaming' (i.e., punishing) persons with mental illness for the commission of crimes (the defense of insanity or nonresponsibility), early research on deception concentrated upon certain aspects of mental illness (Geller, Erlen, Kaye, & Fisher, 1990). Still, feigned mental illness is considered "both uncommon and extremely difficult to sustain" (Anderson, Trethowan, & Kenna, 1956, p. 513). Here, the importance is not so much the uncommon nature of the behavior, particularly in consideration that other research disputes the incidence rate (Flicker, 1956), as it is the malingerer's level of difficulty in maintaining the ruse. Anderson, et al., (1956) refer to a "pull of the reality" (p. 517) as generally reducing one's ability to avoid the fatigue that apparently accompanies malingering. They quote one of their subjects as experiencing fatigue due to the difficulty of maintaining "two processes of thought, one thinking deeply to prevent me from thinking deeply" (p. 517). It is interesting to note that this particular subject was described as behaving in a "childish ridiculous way," and that, in mental testing, her "fatuousness increased to buffoonery" (p. 517). Maintaining a consistent level of malingering appears to have been at least difficult, or impossible.

The fatigue associated with malingering has significant implications for the present study. For example, if a subject is actively and consciously attempting to deceive an examiner, overcoming the fatigue associated with such malingering could be an important part of the success or failure of the ruse. However, in research where the examiner clearly instructs the examinee to malinger, the examinee may not experience the level of heightened tension noted in earlier studies (Goldstein, 1923; Langfeld, 1921). This lower level of tension is experienced because the examinee is not under the pressure to disguise the malingering behavior from the examiner. Therefore, it seems logical to presume a reduction in tension would result from the elimination of conscious deception on the part of the examinee, since both the examiner and the examinee are conscious of malingering in the examinee's responses. If this is so, then it may be true that the examinee would be less likely to experience fatigue. In support of this argument, Anderson, et al., (1956) found that, of 18 subjects in the "simulant group," with very few exceptions, none were able to consistently sustain the simulation to any great extent, and only two feigned "feeblemindedness."

Feigning and the Criminal Justice System

Szasz (1956) considered the relationship of malingering to criminality; however, he viewed the two as being "illusions" of psychopathology, and that psychiatric professionals "substitute the vague and all-inclusive notion of 'mental illness' for all sorts of other problems" (p. 438) unrelated to diagnosis. Wertham (1949) studied characteristics of psychopathology as related to murder and murderers. Identifying two examples of malingered mental illness, Wertham pointedly states that:

> There is a strange, entirely unfounded, superstition even among psychiatrists that if a man simulates insanity there must be something mentally wrong with him in the first place. As if a sane man would not grasp at any straw if his life were endangered by the electric chair. (p. 49)

In examining cases of suspected malingering, Wertham's experience with criminals had taught him to look for a serious charge as a probable cause. He labeled a subject (F.) as malingering, and indicated that such cases were more appropriately handled by the courts than by the psychiatric community. Listing F.'s past history of exhibitionism, overt homosexuality, and the brutal murder of a 3-month old nephew, Wertham described him as "a personality of unheard-of moral callousness" (p. 210; reported in Szasz, 1956, p. 440). One might question Wertham's conclusion that F. was 'malingering.' Moreover, the question of

who should or should not be considered legally responsible for the commission of a crime (the "ultimate question") is not within the role of an expert witness (ABA, 1989).

Whether malingering constitutes mental illness is still open to debate, but Bleuler (1924, in Szasz, 1956) maintained that the determination of the existence of malingering does not necessarily prove that the malingerer is mentally sound and, therefore, responsible for his actions. A malingerer may be faking symptoms of certain characteristics, and yet still be insane.

More recently, studies involving individuals convicted of crimes have suggested that it is not easy for these individuals to feign mental illness. Cornell and Hawk (1989) compared 39 individuals convicted of crimes who had been diagnosed as malingerers by six experienced forensic psychologists to 25 genuinely psychotic defendants. They were able to use inconsistent symptomology to identify malingering, finding that malingerers differed from true psychotics on 14 of 24 clinical presentation variables such as formal thought disorder, hallucinations, affect, and various measures of general presentation.

Parwatikar, Holcomb, and Menninger (1985) investigated malingered amnesia in individuals accused of murder. They contended that such individuals who truly experience amnesia would have high levels of agreement on the MMPI scale known as the 'neurotic triad.' The results indicated that individuals convicted of murder who were intoxicated at the time the crime was committed, and who had evidence of neurosis on the MMPI were likely to be truly amnesiac. However, if the accused stated that he or she was intoxicated but did not show signs of neurosis, malingering should be suspected.

In effect, the psychiatric literature suggests that the profession is not in agreement regarding the definition, medicolegal aspects, or diagnostic validity of malingering. Early psychoanalytic investigations may also be theoretically tainted, since such theories are still unproven and related research is often hopelessly flawed. Also, behavioral aspects of simulation (i.e., malingering) indicate that there are certain characteristics that commonly occur, but may not be universal (Resnick, 1984).

Criminal Justice and Mental Retardation

Mental retardation does not, in and of itself, constitute grounds for incompetence to stand trial (Everington, 1987). Competence, by legal definition, is more convoluted and less easily determined than one might think. Grisso (1986b) considered that the ability to fake incompetence could be manifested by the malingering of functional deficits. He stated that:

> ... some individuals may malinger - that is, may fake incompetent behaviors - or may attempt to simulate competent behavior. That is, they may be motivated to manifest functional deficits or strengths in order to maximize the likelihood of certain desirable consequences of a decision about a legal competency or incompetency. When apparent functional abilities or deficits can be attributed to these causes, they take on a different meaning than when they are believed to have been beyond the individual's control. (p. 21)

Such statements call into question whether an individual could successfully fake mental retardation and be declared incompetent to stand trial.

Ellis and Luckasson (1985) and Noble and Conley (1992) identify several important issues concerning people with mental retardation who become involved in the criminal justice system. For example, the issue of competence to stand trial for people with mental retardation is of great importance to the legitimacy of legal proceedings. This issue is receiving increased attention in the research; however, large numbers of criminal defendants who have disabilities remain unidentified (Burr, 1992). Subsequently, defendants having mental retardation who are not identified are not referred for competency evaluations and proper procedural protections (Burr, 1992; Devault & Long, 1988; Mickenberg, 1981; Noble & Conley, 1992). In fact, there is evidence that defendants with mental retardation can go through the entire process of adjudication and punishment without ever being identified as disabled (Noble & Conley, 1992).

One possible explanation for this would be the so-called "cloak of competence," (Edgerton, 1967), in which people with mental retardation will endeavor to hide their disability from others. Likewise, people administering the justice system are not trained to identify disabilities among offenders (Grisso, 1986b; Luckasson, 1992).

DeVault and Long (1988) question the competence of some defendants with mental retardation to offer a confession. They report one case in which a Native American male charged with second-degree murder waived his rights under Miranda and confessed. Despite IQ evidence that he had mental retardation, and various procedural discrepancies (including the lack of an adaptive behavior measure), the jury concluded he was competent and convicted him.

Winick (1985) asserted that one problem with determining competence is the high cost to society

BEST COPY AVAILABLE

of identifying such individuals. Nevertheless, Cooke, Johnston, and Pogany (1973) suggested that the identification of mental retardation in defendants may reduce the likelihood for the imposition of more severe sentences. While the existence of mental retardation certainly does not routinely constitute incompetence to stand trial, such a finding may increase the possibility that appropriate habilitation may occur.

Another problem in this arena is the possibility of not identifying or misdiagnosing mental retardation or dementia, acquired through severe traumatic brain injury. In a study of 14 death row inmates, researchers concluded that all the subjects had suffered brain trauma, and state that many others probably suffer from undiagnosed psychiatric, neurological, and cognitive disorders that might constitute mitigating circumstances (Lewis, Pincus, Bard, & Richardson, 1987; Penry v. Lynaugh, 1989). This supports the findings of Schretlen and Arkowitz (1990) regarding the symptomatic confusion of mental illness and mental retardation.

In capital cases, people with mental retardation may be statutorily excluded from execution. Five states (Georgia, Kentucky, Maryland, New Mexico, and Tennessee) currently have statutes that prohibit the execution of individuals with mental retardation who are convicted of capital crimes (Burr, 1992). The likelihood of other states enacting similar legislation may be increased by the United States Supreme Court's decision in Penry v. Lynaugh (1989). This decision acknowledged public opinion polls indicating disfavor with proposed executions for criminals who have mental retardation (Gallup, 1989, reported in Burr, 1992). However, the Supreme Court indicated the need for a growing consensus before they would declare that execution of people with mental retardation violated the Eighth Amendment as cruel and unusual punishment.

Given this, several issues arise: It may be at least theoretically possible that defendants who are not mentally retarded may attempt to feign retardation in order to receive a less severe sentence. However, equally true, defendants who do have mental retardation may be mistakenly judged to be feigning, particularly on tests of intelligence. Also, defendants with mental retardation theoretically may try to fake a lower score in order to increase their chances of receiving leniency. Similarly, however, defendants with mental retardation may attempt to hide their disability in order to appear 'normal' (Edgerton, 1967; Resnick, 1984).

McGarry (1986) pointed out that while malingering may be exposed, it takes the trained eye of a professional to discover the deception. In his experience, most cases, including those where the malingerer had studied psychiatric books to perfect the performance, often present "clinical data [that] are histrionic, variegated, and inconsistent and require multiple diagnostic labels in contradistinction to the principle of diagnostic parsimony in good medical practice" (p. 84). In short, defendants malingering mental illness will, usually, overact. McGarry suggests, however, enlisting nurses, correctional officers, attendants, and other nonprofessionals, and emphasizes that their observations of normal behavior while not in the company of the examiner may in fact be admissible in court. This was corroborated by Resnick (1984), who noted that even hospitalized psychiatric patients can sometimes identify malingerers of mental illness.

Despite scoring standards, and given the interpretive nature of testing instruments, the validity and reliability of intelligence tests are important factors necessary to the determination of their use. While several test instruments are normalized on large samples, scores obtained may not be an accurate reflection of the examinee's actual level of competence. Sattler (1988) states,

> Intelligence tests have been criticized because IQ does not adequately relate to many measures of everyday functioning. But can any one measure be expected to correlate highly with behaviors that are multidetermined? Individuals with the same IQ vary widely in their social competence, as well as in the expression of their talents. (p. 77)

IQ tests are not infallible, but they offer information that may be invaluable to the examiner who is trained to identify certain characteristics, including simulation (Noble & Conley, 1992; Resnick, 1984; Sattler, 1988; Schretlen & Arkowitz, 1990).

Summary

There are two major factors that may be used to outline the purpose and rationale for this study. First, despite years of research often closely related to this topic, there is still an acknowledged lack of research in the area of malingered mental retardation (Ogloff, 1990). The vast majority of the research has concentrated upon falsification of mental illness and related disorders on personality and projective tests, not on intelligence tests. Also, the psychiatric interview as a method of determining malingering has been strongly questioned (Rosenham, 1972).

As previously stated, faking mental retardation for purposes of individual gain may be possible. However, given confirming validity data on tests of intelligence, it may be possible to validate a diagnosis of mental retardation through obtaining high levels of agreement in scores and profiles of two accepted, individualized tests of intelligence.

There is a need for awareness within the criminal justice system that high concurrent validity standards may preclude suspicions of malingering on separate tests, unless the examinee does not achieve similar results in both the profiles and the scores achieved. The key to this rests in the validity standards achieved for each test. Several tests do not carry such high validity standards. This is particularly true for group administered tests, tests with poorly or inadequately conducted standardization research, and tests utilizing a "pencil and paper" format (Anastasi, 1988; Jensen, 1980; Sattler, 1988; Swanson & Watson, 1989; Thorndike, et al., 1986a). However, both tests utilized in this study (WAIS-R and SB:FE) have shown good concurrent validity.

Objectives and Implications of this Study

Defendants who have mental retardation are likely to attempt to avoid the mental retardation label for various reasons. The stigma of mental retardation is strong. Individuals whose functioning is mildly impaired may attempt to hide their disabilities. Such individuals have a significant handicap nevertheless. These people may well attempt to 'pass' as nonhandicapped under a "cloak of competence" (Edgerton, 1967). It has been suggested, however, that people of low intellect may also attempt to appear less intelligent in testing situations (Flicker, 1956). Still other individuals in the criminal justice system may see a benefit in and attempt to feign mental retardation in order to reduce the severity of their sentences.

This study demonstrates that corresponding subtests of two highly accepted tests of intelligence are sufficiently valid to accurately identify those individuals who were simulating their responses. When subjects were requested to act like they would if they were trying to convince the examiner that they had mental retardation, they were unable to do so successfully. This study also suggests that the validity standards of the subtests administered in this study are adequate to the extent that they will identify people attempting to simulate mental retardation from those people who have mental retardation.

Additionally, the possibility that individuals may take longer to simulate their responses to test items needs investigation. Early research in this area has suggested this possibility (Goldstein, 1923), but more recent research has not investigated this topic in simulating mental retardation. Individual examinees may also be able to shed some light on the process used in determining simulated answers.

The majority of research in the area of "faking on tests" has concentrated on the examinee's ability to falsify responses to indicate the existence of mental illness. There is a lack of research investigating the possibility that individual examinees could simulate their responses on two different tests of intelligence, thereby providing proof of mental retardation by achieving statistically similar results. If a person could simulate mental retardation on two major tests of intelligence, and achieve similar results, the validity of both tests in determining the existence of mental retardation would be highly questionable. This study was designed to determine whether individuals who have little or no experience in psychological testing could consistently feign mental retardation in answering specific test items on two standard intelligence tests: WAIS-R and SB:FE.

METHOD

The current research study was developed to investigate the possibility of consistently simulating mental retardation across two subtests of two major tests of intelligence. A control (genuine) administration was also administered. In addition, response latency times were measured to examine proposed differences between genuine (control) and simulation (experimental) conditions. Finally, subjects were surveyed for their attitudes and opinions in debriefings after the testing sessions.

Research Questions

The first research question of this study stated: Given Comprehension subtests of the WAIS-R and the SB:FE, was it possible for subjects to simulate mental retardation and obtain normal curve equivalency (NCE) scores that were not statistically different? The second research question of this study stated: Was it possible for subjects to give genuine answers to items from both subtests and obtain NCE scores that were statistically different? The third research question of this study stated: Did it take longer for subjects to simulate mental retardation in their responses to test items than to give genuine responses? Finally, the fourth research question of this study stated: Did subjects find it more difficult to simulate mental retardation than to give genuine responses?

Sample

The subjects of this study consisted of a solicited group of 21 adult Caucasian males, between 20 and 30 years of age (mean age = 25.95 years), and largely comprised of students and staff employees of the University of New Mexico. All subjects were high school graduates, and mean education in years was computed at 15.1 years. These subjects were obtained through the distribution of flyers on the university campus. These flyers offered a small remuneration for the time spent on the experiment ($5.00). For the purposes of this study, only the resulting scores and response time factors were considered to be dependent variables.

Instrumentation

Subjects were administered two corresponding subtests, both titled "Comprehension" from both the WAIS-R (Wechsler, 1981) and the SB:FE (Thorndike, Hagen, and Sattler, 1986a). These subtests evaluate various attributes of the examinee, including verbal comprehension, social judgment, common sense, practical judgment, knowledge of conventional standards, ability to evaluate past experience, and moral/ethical judgment (Sattler, 1988). The questions that comprise the Comprehension subtests of both the WAIS-R (N=16) and the SB:FE (N=42) are stated similarly, as single sentences, i.e., 'What would you do if you were lost in the forest in the daytime?' (WAIS-R) and 'Why do we have fire drills?' (SB:FE) (see Appendix B).

Responses are scored differently according to which test one uses. On the WAIS-R, examinee responses obtain a two, one, or zero score for good, fair, or poor quality, respectively. On the SB:FE, responses are scored one or zero for a good or poor response, respectively. Scoring standards and query responses are outlined in the Administration and Scoring Manuals of each test. These scoring differences constitute an important difference between these tests, and likely would confound each subject's attempt to consistently falsify responses at the same level. In this study, calculated coefficients that determine the amount of agreement between corresponding subtest scores in both administrations denote the degree of concurrent validity between subtests in both conditions.

The internal consistency of the SB:FE's Composite Score (which corresponds to a global IQ score) has been estimated at between .95 and .99 using the Kuder-Richardson technique. The median Composite Score reliability was reported as .97 (Sattler, 1988). The median reliability for the Comprehension subtest was determined to be .89 (Thorndike, et al., 1986). The mean internal consistency of the WAIS-R Full-Scale IQ was reported at .88 for all age groups, and reliability for the Comprehension subtest was calculated at .84, using a split-half correlation with a Spearman-Brown correction (Wechsler, 1981).

Concurrent validity research comparing the global scores of the SB:FE and the WAIS-R estimates the correlation at between .85 and .91 (Sattler, 1988; Thorndike, et al., 1986), indicating a high degree of validity. The Comprehension subtests, however, appear to load differently on the general factor of intelligence. The WAIS-R Comprehension subtest has a median loading of .78 (good), while the SB:FE Comprehension subtest has a median loading of .68 (moderate). Both subtests have high loadings on the verbal comprehension factor (Sattler, 1988). Other forms of validity (face, content, and construct) appear to

be adequate for both instruments (Anastasi, 1988).

Data Collection Procedures and Analysis

Two subtests were administered twice to each subject, taking approximately one hour. In the control condition, ten subjects were given the genuine administration first. In the experimental condition, eleven subjects were given instructions to simulate their responses to appear mentally retarded on the first administration. The process, intended to control for possible administration order effects, was then reversed for each subject. To reduce possible order effects, administration of the two conditions were randomly ordered using a standard table of random numbers; odd numbers being control condition first, even numbers being experimental condition first. This procedure was implemented to control for possible advantages and/or disadvantages to the order of each condition's administration (Borg & Gall, 1982). All testing sessions were recorded on videotape to improve scoring and timing accuracy.

Neither the WAIS-R or the SB:FE offer specific instructions for the Comprehension subtests. The examiner began the control-first administration by saying "I am going to ask you some questions," prior to beginning these subtests. When subjects were administered the experimental (simulating) condition, the following instructions were read to them:

> I want to see how well you can act like you have mental retardation. You will be given two tests that will test your ability to comprehend certain social situations. You should try to respond in a way that would identify you to me as a person with mental retardation. You must try to respond to each of the tests using the same level of retardation. Now, remember, you should perform like you would if you were trying to convince me that you have mental retardation.

Two forms of data were collected in this study, quantitative and qualitative. Quantitative data consisted of results from two separate measures: Equally calibrated standardized scores of subtest results under both administration conditions and, measured time lapses between item presentation and responses given the two different administration conditions, typically called response latency (Alberto & Troutman, 1986). The first data measure (standardized scored responses) was expected to indicate that subjects produce similar scores when comparing the two control administrations, but different scores when comparing the two experimental administrations. The second data measure (response latency) was expected to indicate that subjects require significantly more time to develop an appropriate response when simulating mental retardation than in the control, or genuine, condition.

Each subject's responses were scored in accordance with each test's specific scoring procedures. Scored responses for both subtests were calculated first as raw scores, then translated to standard scores using the tables supplied in the test manuals. Since the reported standard score tables on each subtest are calibrated differently, all standard scores were then transformed to corresponding percentiles, and the percentiles were transformed again to fit the normal curve equivalency (NCE). This resulted in equally calibrated standard scores based upon a scale with a mean of 50 and a standard deviation of 10.

Time latency of responses, beginning at the end of each question posed by the examiner and ending at the start of the examinee's scored response, were measured by viewing the videotape of each subject and using a stopwatch. Each subject's response time during separate conditions was recorded, averaged, and compared for potential differences in time required for response formulation in the two conditions. Interobserver reliability for the time factor was determined to be 1.0, indicating no discrepancy in the measure of latency.

Finally, qualitative data were obtained in a short debriefing session following both the testing conditions. On videotape, subjects were asked to answer several questions regarding their thoughts and opinions about the different testing conditions and their subsequent responses, and how they developed responses to each condition (see Table 9). These qualitative responses were then broadly analyzed to indicate similarity in subjects' overall perception of different conditions and related opinions, thoughts, and ideas. These qualitative data are reported in percentages of subjects making identical or very similar observations.

Table 9. Qualitative Debriefing Questions.

1. Give me your impressions of what you have just done.
2. Did you find it easy or difficult to fake retardation?
3. What questions did you find more difficult to fake? Why?

BEST COPY AVAILABLE

Table 9. Qualitative Debriefing Questions (continued).

4. Did you have to think harder when you were faking retardation or when you were giving truthful answers?
5. How did you go about determining the kinds of answers you gave when you were faking retardation?

As stated, these two tests have different subtest scoring calibrations, so it was necessary to use a method of equalizing the resulting subscale scores. Standardized normal curve equivalency (NCE) scores for corresponding subtests on both the control and the experimental conditions were compared for statistical difference using two methods; separate one-way repeated measure analysis of variance (ANOVA), and a Wilcoxon nonparametric comparison. Time latency of responses was analyzed to compare potential differences in time required for response formulation for the two conditions. Again, a one-factor repeated measure analysis of variance (ANOVA) was used to compare for statistical differences between conditions. Simple descriptive statistics were also determined and reported.

The first and second research questions were examined using two, one-factor repeated measure ANOVAs to compare NCE scores on both the simulated retardation administrations and both the genuine administrations of the Comprehension subtests of the WAIS-R and the SB:FE. This was to determine if subjects achieved scores that were or were not statistically different on administrations of each corresponding subtest under the experimental or control conditions. It was expected that the experimental (simulated) administration would not yield similar results, but the control (genuine) administration would yield similar results. Also, a Wilcoxon nonparametric test was utilized in effort to control for the probability of less than normal distribution within the data.

The third research question was examined using two separate repeated measure ANOVAs to compare individual examinee response latency periods. These analyses differed from the NCE analyses, as this comparison investigated latency times for questions on the same subtest under different conditions, rather than comparing NCE scores for different subtests under the same conditions. Results determined if latency times in answering questions on the same subtests in the experimental and control conditions were significantly different. The fourth research question of this study, did individual examinees find it more difficult to simulate mental retardation on the subtests administered than to give genuine responses, was determined using percentage of agreement observed in the qualitative data obtained during debriefing sessions.

RESULTS

Each subject was administered the Comprehension subtests of the WAIS-R and the SB:FE. Under the control condition, each subject responded genuinely to the test questions. In the experimental condition, each subject was asked to respond to the test questions with the intention of convincing the examiner that he was mentally retarded. For the purpose of subsequent statistical analysis, subscale raw scores were converted to normal curve equivalency (NCE) scores (mean = 50). NCE scores for each subject were calculated across both conditions and for both tests (see Table 10). Summary data, including average scores for the subject population, across both conditions and tests, are reported in Table 10.

Visual examination of Table 10 indicates that in the control (genuine) condition, scores were relatively consistent. However, in the experimental (simulated) condition, the scores are somewhat inconsistent, with only two subjects able to achieve similar scores in the range of mental retardation (Subjects 1 & 12). Finally, the NCE scores for the experimental condition indicate that some of the respondents interpreted mental retardation as meaning that all questions would be answered incorrectly. These examinees obtained minimal NCE scores (i.e., 1) on both tests during the experimental condition.

The mean NCE scores for each condition (as reported in Table 11) indicate that the subjects' performance under control conditions were consistent, despite the wide range of scores achieved. However, mean NCE scores under experimental conditions indicate a wide discrepancy in achieved scores. In addition, the standard deviations for the experimental condition of both tests indicate a wide variance in NCE scores, while control administration standard deviations varied considerably less.

Table 10. Normal Curve Equivalency Data by Individual Subject.

Subject	WAIS-R Cont.	WAIS-R Exp.	SB:FE Cont.	SB:FE Exp.
1	36	15	31	13
2	64	1	34	1
3	43	36	60	1
4	57	15	60	1
5	57	50	60	68
6	50	1	60	1
7	99	50	76	55
8	50	22	76	13
9	57	22	68	1
10	57	1	60	1
11	57	1	55	1
12	57	15	76	15
13	50	1	60	1
14	57	1	55	1
15	78	1	60	1
16	85	7	76	1
17	71	64	55	1
18	78	15	68	1
19	78	29	76	20
20	57	15	76	1
21	78	57	55	55

Table 11. Descriptive Statistics - WAIS-R and SB:FE - NCE scores

	Mean	SD	SE	Minimum	Maximum
WAIS-R Control	62.67	15.28	3.33	36	99
WAIS-R Experimental	19.95	20.34	4.44	1	64
SB:FE Control	61.76	12.68	2.77	31	76
SB:FE Experimental	12.05	20.74	4.53	1	68

The first research question of this study asked whether individual examinees can simulate retardation on separate, corresponding subtests of the WAIS-R and the SB:FE and achieve scores that are not statistically different. A one-way repeated measures analysis of variance (ANOVA) of the experimental condition NCE scores indicated that subjects' responses to each individual subtest were significantly skewed, F (1, 20) = 4.801, p<.04. While certain subjects appeared able to score lower in the experimental condition, they could not consistently achieve statistically similar NCE scores on both Comprehension subtests. Mean scores were also significantly different (see Table 12 and Figure 1).

The second research question of this study asked whether or not subjects could obtain statistically similar results on both Comprehension subtests of the WAIS-R and the SB:FE under control (genuine) conditions. A visual inspection of Table 11 and Figure 1, and results of a one-way repeated measure ANOVA of NCE scores obtained under control conditions (see Table 12) indicate that subjects did score similarly on both tests, F (1, 20) = 0.072, p<.79. There was no significant difference between the two subtest administrations under control conditions.

It cannot be assumed that the distribution on either condition, particularly the experimental condition, is representative of normal distribution. In effort to control for this, a Wilcoxon nonparametric test was applied. With this statistic, scores that are tied (exact duplicates) are eliminated as a control measure. Results indicated that, in the control condition, subjects did not significantly differ in their scores (Z = -.052 [no eliminations]). However, in the experimental condition, subjects' responses were skewed significantly (Z = -2.27 [8 eliminations]). This strongly supports the results of the ANOVAs used to examine the first two research questions.

The possibility of an administration order effect for each condition was also examined. Results of a two-factor ANOVA of NCE scores for each condition and administration order indicate that there was no order effect, F (1, 19) = 1.458, p<.242. In addition, there was no interactive order effect by treatment, F (3, 19) = 1.243, p<.278.

Table 12. One factor ANOVA of NCE Scores for each condition.

	Mean Diff.	F-Ratio	p value
WAIS-R v. SB:FE Control	.905	0.07	.79
WAIS-R v. SB:FE Experimental	7.905	4.80	.04

The third research question of this study asks whether individual examinee response latency periods in answering questions differ between the experimental (simulated) or control (genuine) conditions for a given test. Response latency scores for each subject were averaged for each condition (see Table 13). Descriptive statistics for response latency means by test and condition are reported in Table 14. Individual subject response latency times for both administrations of the WAIS-R and SB:FE are presented in Figures 2 and 3, respectively. Subjects' mean response latency for each condition indicate that the experimental condition showed longer periods of response formulation than the control condition for almost every subject.

For the WAIS-R, the mean latency period (MLP) for the control condition was 2.75 seconds, while for the experimental condition the MLP was 4.58 seconds. In a one-way, repeated measure ANOVA, the difference between these two means was statistically significant, F (1, 20) = 23.41, p<.0001. Second, for the SB:FE, the MLP for the control condition was 1.49 seconds, while for the experimental condition the MLP was 3.62 seconds. In a one-way, repeated measure ANOVA, the difference between these two means was statistically significant, F (1, 20) = 73.47, p<.0001. Once again, the standard deviations of experimental condition MLPs were larger than those of the control condition, albeit less so than in the NCE scores. Overall, these results provide strong support for the concept that it takes longer for an examinee to develop a simulated response than a genuine response (see Table 14). Visual examination of Table 13 further suggest that at least certain questions in the WAIS-R require longer periods of response formulation than the SB:FE for both conditions. However, this may be a factor of the number of items in each subtest, or just subject fatigue.

Table 13. Mean Response Latency by Subject and Condition (in seconds).

Subject	WAIS-R Cont.	WAIS-R Exp.	SB:FE Cont.	SB:FE Exp.
1	4.00	3.12	1.98	2.31
2	1.81	2.94	1.10	2.83
3	4.06	3.38	1.38	2.26
4	6.00	10.81	2.36	6.45
5	2.50	3.00	1.48	1.95
6	2.88	4.69	1.07	3.52
7	1.62	3.69	1.43	2.12
8	1.81	3.75	1.74	3.26
9	2.81	4.06	1.24	4.21
10	2.12	4.75	1.48	5.40
11	4.50	4.81	2.24	4.43
12	2.38	6.94	1.90	4.69
13	2.44	4.44	0.88	2.81
14	3.00	4.38	1.64	4.21
15	2.12	2.00	1.12	2.40
16	1.69	6.19	0.90	4.74
17	1.44	3.06	0.83	2.86
18	2.12	4.81	1.36	4.71
19	2.06	5.31	1.52	4.50

20	3.31	2.88	1.55	2.74
21	3.00	7.19	2.10	3.55

The fourth research question of this study was concerned with whether or not the examinees found it more difficult to simulate mental retardation on the subtests administered than to give genuine responses. This question was assessed by examining the verbal responses obtained during debriefing sessions that followed each testing session. Each subject was asked to answer five questions related to the activity in which they had participated. Responses were transcribed, compared for similarity of response, and percentages of agreement were tabulated (see Table 15).

Results indicated that 67% of the examinees found it difficult to simulate retardation, and 33% found it easy to simulate retardation. However, when asked if simulating mental retardation or giving genuine responses required more thought, 95% of subjects (20/21) stated that simulating mental retardation required more thought. While these self-report data are obviously subjective, the high agreement in question 2 would suggest that subjects found it more difficult to continue simulating than to give genuine answers, despite the 33% who stated they did not find it difficult to simulate mental retardation (see Appendix B).

Table 14. Descriptive Statistics - WAIS-R and SB:FE - Response Latency

	Mean	SD	SE
WAIS-R Control	2.75	1.12	0.25
SB:FE Control	1.49	0.44	0.10
WAIS-R Experimental	4.58	1.96	0.43
SB:FE Experimental	3.62	1.22	0.27

Table 15. Percentage of Agreement to Debriefing Questions.

1. Did you find it easy or difficult to fake retardation?
   - Difficult 67% (14 out of 21)
   - Easy 33% (7 out of 21)
2. Did you have to think harder when when you were faking retardation or giving truthful answers?
   - Faking mental retardation 95% (20 out of 21)
   - Telling Truth 5% (1 out of 20)

Interrater reliability was determined through the cooperation of College of Charleston colleagues; one trained as a school psychologist, the other a graduate assistant. Reliability on individual item scoring was estimated at .94 and .96 on viewing two, randomly selected subjects' performance on both treatment conditions. Reliability on latency periods was estimated at 1.0 (see Appendix C).

IQ and the Death Penalty: Validating Mental Retardation

A Paper presented to the 101st Annual Conference of the American Psychological Association

Denis William Keyes, Ph.D.

INTRODUCTION 1
Reliability and Validity of Test Instruments 1
Response Falsification 2
The Criminal Justice System 2
Statement of the Problem 2
Purpose and Rationale of the Study 3
REVIEW OF THE LITERATURE 4
Test Validity 4
Validity: WAIS-R & SB:FE 7
Malingering and Deception 8
Defining the Concept 8
Characteristics of Malingering 10
Assessment of Feigning 12
Early Research - Deception and Time Latency 13
Feigning Mental Retardation 15
Difficulties in Feigning Mental Illness 18
Criminal Justice and Mental Retardation 19
Summary 20
Objectives and Implications of this Study 21
METHOD 22
Research Questions 22
Sample 22
Instrumentation 22
Data Collection Procedures and Analysis 23
RESULTS 24
DISCUSSION 28
Limitations 29
Implications of the Current Study 29
Directions for Future Research 30
Conclusions 31
REFERENCES 34

DISCUSSION

The present research study was designed to examine the possibility of successfully simulating on two corresponding subtests from two major test batteries of intelligence, the SB:FE and the WAIS-R. One important factor of this topic is based upon legal implications that may be considered when a criminal defendant has mental retardation. For example, jury instructions for deliberations must include mental retardation as a mitigating factor (Penry v. Lynaugh, 1989). However, if mental retardation can be successfully simulated, sentences may be incorrectly based upon improper diagnosis of mental retardation.

The accurate determination of one's functioning level is paramount to identification of the existence of mental retardation. The main rationale for this study was to investigate the possibility that the validity standards of both the WAIS-R and the SB:FE would make it impossible to simulate mental retardation consistently at the same level. If one could simulate having mental retardation to such a degree that its existence is similarly validated on two measures of intelligence, then the validity of both the WAIS-R and the SB:FE would be seriously questioned. However, if it is not possible to simulate mental retardation at similar levels on both tests, then the validity of both the WAIS-R and the SB:FE is not only intact, but enhanced. Moreover, defendants who have been accused of simulating on both tests, but achieved statistically similar results, may have an important recourse based upon the findings in this study.

The findings in the present study suggest that individuals cannot simulate mental retardation consistently at the same level on both tests of Comprehension, but that when they are giving genuine responses, their resulting scores will not be significantly different. These findings have important implications regarding the existence of mental retardation in those who are involved with the criminal justice system. If test validation standards may be used to certify the existence of mental retardation, then those people who are simulating mental retardation may be more accurately exposed, and those who are not simulating may receive a more just sentence from identifying the seriousness of their disability.

Another major issue in this study was to consider the question of time latency in the simulating test item responses. This factor received considerable investigation in early research on simulating (Goldstein, 1923; Langfeld, 1921; Marston, 1920) but has been neglected in the more recent literature. Questions dealing with the 'types' of personality of the subjects (Goldstein, 1923) were discarded, in favor of strictly comparing the latency of genuine responding on test items to the latency of simulating in responding to test items. Results of this study suggest that latency periods when responding to test items were significantly longer when simulating than when giving genuine responses. This finding has limited use, of course, since the examiner will not know, initially, if the examinee is simulating. However, if longer response times were noted during test sessions, particularly when resulting scores are compared between two tests are significantly different, then the suspected existence of simulating may be more accurately determined. This is even more important when the examinee has an opportunity for secondary gain by simulating, the very definition of the term 'malingering.'

The findings of this study enhance both the concurrent validity and the construct validity of both the subtests utilized. The lack of an order effect, and the lack of any interactive order effect suggests that both of these tests investigate similar abilities of the subjects being tested. Additionally, the similarity of each subject's scores on the control condition suggests that the tests are concurrently valid.

Finally, qualitative data collected have important implications to the understanding of processes involved when people are attempting to answer falsely in responding to test items. All but one of the examinees found it considerably more difficult to simulate mental retardation than to give genuine responses. This supports the findings of increased latency periods when simulating retardation in responding to test items, suggesting that it takes longer to develop false answers in responding because it is more difficult to come up with an incorrect, simulated response than a genuine one.

Summary

The method used to investigate the research questions of this study involved the use of 21 subjects who were tested on two corresponding subtests of two major intelligence tests. Each subject was asked to give both genuine and simulated responses to each subtest, in two administrations of each subtest. The order of treatment administration was randomly assigned, and no order effect was determined. Each testing session was videotaped to improve the accuracy both of latency calculation and scoring of responses.

The results of the data analyses strongly support the two major concepts investigated within this study. Individual subjects did not appear able to simulate mental retardation consistently on both subtests and achieve statistically similar results, and latency times for simulating appear to be significantly longer than latency times for genuine responding. The findings also suggest that genuine responding will yield similar scores on both subtests, and qualitative data suggest that it is more difficult to simulate retardation

BEST COPY AVAILABLE

on a test than to give genuine answers.

Limitations

There are several limitations to the application and generalization of this study. The subjects involved in this project presumably did not have the same motivation to simulate as someone who would gain by simulating mental retardation (i.e., avoiding death penalty, or better prospects for a habilitation program in prison). While minor financial remuneration was given to each subject, there was no contingency of increased reward dependent upon results of the testing (Heaton, Smith, Lehman, & Vogt, 1978).

Generalization may be limited because the subjects of this study were 21 white male, high school graduates. Current statistics suggest that while the majority of the incarcerated individuals in this country are white, the percentage of nonwhite (black and Hispanic) inmates exceeds their actual representation in the general population (Noble & Conley, 1992). Additionally, had this study made use of defendants who were incarcerated, it would have raised potentially serious ethical questions about whether such testing would give them the idea of simulating retardation, and increase their awareness of opportunity for different treatment in the criminal justice system. Making such information public might raise further ethical problems as well, since there is an important element of confidentiality to the use of IQ tests, particularly within institutional systems that are known to make use of such tests, i.e., prisons. Giving people exposure to the tests for research purposes rather than for use in medical and/or psychiatric situations may not be considered legitimate under such circumstances.

The possible parallels drawn to the criminal justice system are also limited by the fact that this study made use of only one subtest from each battery, rather than a full scale assessment. The amount of time required, and the cost of materials for such a study prohibited the use of the WAIS-R Full Scale IQ and the SB:FE Composite Score.

It should also be remembered that the subtests utilized in this study are from two intelligence tests with exceptionally high validity standards. It is unlikely that results of intelligence tests with less impressive validity standards (i.e., Revised-Beta [ ]) could be utilized in any subsequent replicating research.

As with so much other research, the small sample size and demographics of the sample used also call into question the generalizability of the findings. As stated, white males do not represent all people in prison for whom this study might apply. The cultural fairness of the use of IQ tests with members of various minority groups has long been questioned in the literature (Mercer, 1973), and in litigation (Larry P. v. Riles, 1984; PASE v. Hannon, 1980). Also, the tests used herein are based upon a theory of intelligence the construct validity of which has been challenged (Jensen, 1980; Wainer & Braun, 1988).

Certain examinees obtained minimal NCE scores (i.e., 1) on both tests during the experimental condition. This does not indicate response consistency however, since several of these examinees answered all subtest questions incorrectly on the experimental condition. Answering all questions incorrectly may be likely for persons functioning in the more severe levels of mental retardation, but it is probably not realistic for those functioning in the mild to moderate levels. It appears extremely unlikely that a defendant could accurately simulate mental retardation to the extent that resulting scores and profiles would consistently match. In this study, only two subjects were actually able to achieve similar scores in the experimental condition (subjects 1 and 12). Responding incorrectly to all test items, however, indicates that the person attempting to simulate mental retardation does not have a clear or accurate understanding of the condition. Such individuals are easily identified as attempting to simulate the condition.

Finally, generalizability of the qualitative data may limit the findings of this study. Qualitative data reported here are self-reported, and as such, quite possibly lack validity. Test-retest reliability on such data has often illustrated serious discrepancies between varying opinions and attitudes at different times (Jahoda & Warren, 1966). Actual analysis of qualitative responses requires much more investigation than reporting percentiles of agreement. For example, the responses of individual subjects to questions posed in the debriefing reflect considerable variability to how they perceived mental retardation (see Appendix B). Also, the amount of exposure that each individual subject had to people with mental retardation would quite likely affect their performance on these subtests. An example of this factor is evident in the NCE scores of both subjects 2 and 3: Having had no experience interacting with people who have mental retardation, they answered every question wrong in the experimental treatment condition.

Implications of the Current Study

The findings of this study do have some important implications. Results suggest that individuals cannot obtain consistent scores on valid tests of intelligence and successfully simulate mental retardation. The consistency in the control results suggests that the study has importance to the determination of each

subtests' validity.

The implications for participants in the criminal justice system may be significant, particularly for defendants accused of simulating on tests. If the subjects could not consistently simulate mental retardation on only two subtests, then one must question if it is logical to conclude that one individual examinee could consistently simulate mental retardation on two entire test batteries.

Findings related to the third research question suggest that simulating may take longer than giving genuine responses. Earlier research (Langfeld, 1921; Marston, 1920) proposed this conclusion, and the present research supports this idea. This factor may prove to be a vital one in the determination of simulation in intelligence testing.

The reliability of the findings in this study also has important implications. Despite strict standardization standards, interrater reliability data collected indicate some minor disagreements in scoring the responses both in the control and experimental conditions (see Appendix D). This was true for both conditions, and suggests that individual examiners may disagree on the quality of individual examinee's responses. Still, the overall reliability of both instruments utilized in this study has proven impressive over the years (Anastasi, 1988; Sattler, 1988), and the findings in this study appear to support their validity.

The examination of the first and second research questions strongly support the concurrent validity of the Comprehension subtests of both instruments. Results of both control treatments indicate that when subjects answered genuinely, there was no significant difference in their NCE scores. The previous research on these two instruments is supported by these findings (Thorndike, Hagen, & Sattler, 1986b; Wechsler, 1981). The examination of order of treatment administration had no significant effect, and that there was no interactive effect as a result of prior exposure to either test. This is a potentially confounding issue, and it is important to consider that, even with prior exposure to the test items, only two examinees were able to simulate their responses consistently (but only by answering every question incorrectly).

The examination of the fourth research question considered the importance of qualitative data collected in post-testing debriefing sessions. The determination of whether examinees found it more difficult to simulate retardation than to give genuine responses holds significant implications to this study. Generalizing the results to the validity of the tests may be important to future research in this area, specifically whether examinees require more time and must think more to simulate. As subject eight suggested, "The hardest part was trying to maintain a certain level of mental retardation." Also, subject nine stated that, "The answer popped into my head, and I had to stop and think of an answer that someone who was mentally retarded would give." This supports earlier research that suggested examinees had to maintain two levels of thought; one for their 'normal' selves, and one to maintain the ability to simulate (Anderson, et al., 1956).

Directions for Future Research

The rejection of both null hypotheses suggests that further research in this area is necessary. As previously suggested, future research should concentrate on a broader scope of study, including an increase in the number of sample subjects. The demographic characteristics of the subjects included in future research should be more representative of the population at large. This is especially significant in light of the demographics of the population within the criminal justice system.

Instrumentation utilized in the current study was chosen based upon the high level of concurrent validity of the tests and the similarity between subtests. As the results suggest, both Comprehension subtests appear to test very similar characteristics of the subjects. Future research should attempt to expand the scope of instrumentation by providing an increase in either the number of subtests administered, or perhaps administration of the entire test batteries.

Future investigation of the possibility of consistently simulating mental retardation should provide increased motivation and incentives for simulating more accurately (Heaton, et al., 1978). Subjects who perceive a better 'pay-off' for a better 'performance' may become more involved in the attempted simulation of mental retardation. This may better simulate the situation that true malingerers experience, that being the potential for secondary gain when successful in simulating.

A comparison of the latency of responding between the different conditions presented in this study may have much more effect if future research investigated the latency of responding observed in a sample of individuals already diagnosed with mental retardation. Persons with mental retardation characteristically appear to be somewhat more impulsive (Patton, Beirne-Smith, & Payne, 1990), and if analyses of their responses to test items suggest this impulsivity, the latency factor becomes a stronger indicator for the existence of malingering.

The amount of prior exposure that individual subjects have had to persons with mental retardation

may be an important consideration for future researchers. The current study began with the assumption that many diverse groups of people may consider simulating retardation for different potential rewards. This may, in fact, be true for several other reasons than just those dealing with the criminal justice system. For example, people may consider the potential advantages of receiving public support often available to people who have mental retardation or dementia, such as Supplemental Security Income or other forms of disability insurance. Among those who might choose to attempt this ruse are doubtless people who have had at least some experience with persons who have mental retardation. It may be interesting (and justified) to sample those people who may be likely to simulate mental retardation successfully; namely, incarcerated individuals who have had experience or exposure with persons who have mental retardation.

Conclusions

The lack of research in this particular arena (simulated mental retardation) implies that such topics may have been overlooked in the literature. If allegations of faking arise, valid diagnosis should depend upon more than the results of the administration of a single test of intelligence. Within the criminal justice system, the use of less valid and less reliable tests that are faster and easier to administer has not been atypical, and such practices increase the chances for error to occur. Many such tests do not appear to be as sensitive to measuring specific factors dealing with an individual examinee's functioning level (Anastasi, 1988; Lezak, 1983).

As previously stated, the definition of mental retardation requires a measure of adaptive behavior as well as intelligence, a fact that may well prevent potential misdiagnosis. While one may simulate responses on a test, adaptive scales often require a respondent other than the examinee, thus necessitating the enlistment of a second player in the ruse. Also, a diagnosis of mental retardation is, technically, only possible up to the subject's eighteenth birthday (Grossman, 1983), after which the diagnosis of the condition is referred to as dementia (DSM III-R, 1987). Thus, school records may become vitally important in the case of later identification, for the purpose of determining the prior existence of the condition.

The factor of response latency differences between the control and the experimental conditions supports research long since completed. However, such information should not be forgotten, as the present research suggests that it may still have applicability. While research standards and practices may change, the characteristics of people may not.

The concurrent validity standards of major intelligence tests, particularly those administered in the present study, appear to have potential as one method for exposing simulated mental retardation. The capability of individual examinees to successfully simulate mental retardation is doubtful when results of two standardized tests of intelligence are clearly similar. Similar results on two tests of intelligence may be the best method of confirming mental retardation in individuals accused of faking.

REFERENCES

Albert, S., Fox, H., & Kahn, M. (1980). Faking psychosis on the Rorschach: Can expert judges detect malingering? Journal of Personality Assessment, 44, 115-119.

Alberto, P., & Troutman, A. (1986). Applied behavior analysis for teachers (2nd Ed.). Columbus, OH: Merrill.

American Bar Association. (1989). ABA criminal justice mental health standards. Washington, D.C.: ABA.

American Psychiatric Association (1983). Diagnostic and statistical manual of mental disorders. Washington, D.C.: American Psychiatric Association.

American Psychiatric Association (1987). Diagnostic and statistical manual of mental disorders (revised - 3rd Ed.). Washington, D.C.: American Psychiatric Asso.

Anastasi, A. (1988). Psychological testing (6th ed.). New York: Macmillan.

Anderson, E., Trethowan, W., & Kenna, J. (1956). An experimental approach to the problem of simulation in mental disorder. Proceedings of the Royal Society of Medicine, 49, 513-520.

Bandura, A. (1969). Principles of behavior modification. New York: Holt, Rinehart and Winston.

Barnett, W. (1986). Definition and classification of mental retardation: A reply to Zigler, Balla, and Hodapp. American Journal of Mental Deficiency, 91, 111-116.

Bash, I. & Alpert, M. (1980). The determination of malingering. Annals of the New York Academy of Science, 347, 86-99.

Benton, A. (1945). Rorschach performances of suspected malingerers. Journal of Abnormal Psychology, 40, 94-96.

Benton, A. & Spreen, O. (1961). Visual memory test: The simulation of mental incompetence. Archives of General Psychiatry, 4, 79-83.

Bernard, L. (1990). Prospects for faking believable memory deficits on neuropsychological tests and the use of incentives in simulation research. Journal of Clinical and Experimental Neuropsychology, 12, 715-728.

Bigler, E. (1990). Neuropsychology and malingering: Comment on Faust, Hart, and Guilmette. Journal of Consulting and Clinical Psychology, 58, 244-247.

Bigler, E. & Ehrfurth, J. (1981). The continued inappropriate use of the Bender Visual Motor Gestalt test. Professional Psychology, 12, 562-569.

Bleuler, E. (1924). Textbook of psychiatry. New York: Macmillan.

Borg, W. & Gall, M. (1983). Educational research: An introduction (4th ed.). New York: Longman.

Brody, G., Stoneman, Z., Millar, M., & McCoy, J. (1990). Assessing individual differences. Journal of Personality Assessment, 54, 401-411.

Brown, B. & Courtless, T. (1971). The mentally retarded offender. DHEW Publication No. (HMS) 72-90-39, Washington, D.C.: U.S. Government Printing Office.

Bruhn, A. & Reed, M. (1975). Simulation of brain damage on the Bender Gestalt Test by college subjects. Journal of Personality Assessment, 39, 244-255.

Burr, R. (1992). Introductory essay. In Conley, R., Luckasson, R., & Bouthilet, G. The criminal justice system and mental retardation: Defendants and victims (pp.245-248). Baltimore, MD: Brookes.

Campbell, J. (1941). Psychiatry and military service. Kentucky Military Journal, 39, 110.

Caplan, M., Lubin, B., & Collins, J. (1982). Response manipulation on the Depression Adjective Check List. Journal of Clinical Psychology, 38, 156-159.

Cronbach, L. (1971). Test validation. In R.L. Thorndike, Educational Measurements (2nd ed.) (pp. 443-507). Washington, D.C.: American Council on Education.

Conley, R., Luckasson, R., & Bouthilet, G. (1992). The criminal justice system and mental retardation: Defendants and victims. Baltimore: Brookes.

Cooke, G., Johnston, N., & Pogany, E. (1973). Factors affecting referral to determine competency to stand trial. American Journal of Psychiatry, 130, 870-875.

Cornell, D. & Hawk, G. (1989). Clinical presentation of malingerers diagnosed by experienced forensic psychologists. Law and Human Behavior, 13, 375-383.

Curran, W., McGarry, A., & Shah, S. (Eds.). (1986). Forensic psychiatry and psychology: Perspectives and standards for interdisciplinary practice. Philadephia: F.A.Davis.

Dahlstrom, W., Welsh, G., & Dahlstrom, L. (1975). An MMPI handbook: Vol. 2. Research applications. Minneapolis, MN: University of Minnesota Press.

Daniel, A. & Resnick, P. (1987). Mutism, malingering, and competency to stand trial. Bulletin of the American Academy of Psychiatry and the Law, 15, 301-308.

Davidson, H. (1965). Forensic psychiatry (2nd ed.). New York: The Ronald Press.

Denkowski, G. & Denkowski, K. (1985). The mentally retarded offender in the state prison system: Identification, prevalence, adjustment, and rehabilitation. Criminal Justice and Behavior, 12, 55-70.

DeVault, S. & Long, D. (1988). Adaptive behavior, malingering and competence to waive rights: A case study. American Journal of Forensic Psychology, 6, 3-15.

Diamond, B. (1956). Simulation of sanity. Journal of Social Therapy, 2, 158-165.

Drob, S. & Berger, R. (1987). The determination of malingering: A comprehensive clinical-forensic approach. Journal of Psychiatry and the Law, 15, 519-538.

Edgerton, R. (1967). The cloak of competence: Stigma in the lives of the mentally retarded. Berkeley, CA: University of California Press.

Eisner, D. (1985). The forensic psychologist as expert witness: Satisfying requirements for qualifications, psychological evaluations, and opinions. American Journal of Forensic Psychology, 3, 19-27.

Ellis, J. & Luckasson, R. (1985). Mentally retarded criminal defendants. The George Washington Law Review, 53, 414-493.

Ellis, J. & Luckasson, R. (1992). Brief of American Association on Mental Retardation et al. as amici curiae in support of petitioner in re Johnny Paul Penry, Petitioner, v. James A. Lynaugh, Director, Texas Department of Corrections, Respondent. In Conley, R., Luckasson, R., & Bouthilet, G. The criminal justice system and mental retardation: Defendants and victims (pp.245-278). Baltimore, MD: Brookes.

BEST COPY AVAILABLE

Everington, C. (1987). A Study of the effectiveness of the Competence Assessment for Standing Trial for Defendants with Mental Retardation (CAST-MR) as a measure of competence to stand trial in mentally retarded criminal defendants. Unpublished doctoral dissertation, University of New Mexico, Albuquerque, NM.

Everington, C. & Luckasson, R. (1989). Addressing the needs of the criminal defendant with mental retardation: The special educator as a resource to the criminal justice system. Education and Training in Mental Retardation, 24, 193-200.

Everington, C. & Luckasson, R. (1992). The Competence Assessment for Standing Trial for Defendants with Mental Retardation (CAST-MR). Columbus, OH: I.D.S.

Faust, D. & Guilmette, T. (1990). To say it's not so doesn't prove that it isn't: Research on the detection of malingering: Reply to Bigler. Journal of Clinical Psychology, 58, 248-250.

Faust, D.; Hart, K.; Guilmette, T.; & Arkes, H. (1988). Neuropsychologists' capacity to detect adolescent malingerers. Professional Psychology Research and Practice, 19, 508-515.

Flicker, D. (1956). Malingering - a symptom. Journal of Nervous and Mental Diseases, 123, 23-31.

Geller, J., Erlen, J., Kaye, N., & Fisher, W. (1990). Feigned insanity in nineteenth-century America; Tactics, trials, and truth. Behavioral Sciences and the Law, 8, 3-26.

Gill, M. (1941). Malingering. Bulletin of the Menninger Clinic, 5, 157.

Goebel, R. (1983). Detection of faking on the Halstead-Reitan Neuropsychological Test Battery. Journal of Clinical Psychology, 39, 731-742.

Goldberg, J. & Miller, H. (1986). Performances of psychiatric inpatients and intellectually deficient individuals on a task that assesses the validity of memory complaints. Journal of Clinical Psychology, 42, 792-795.

Goldstein, E. (1923). Reaction times and the consciousness of deception. American Journal of Psychology, 34, 562-581.

Gorman, W. (1982). Defining malingering. Journal of Forensic Sciences, 27, 401-407.

Gorman, W. (1983). Malingering (letter). Journal of the American Medical Association, 250, 31.

Gorman, W. (1984). Neurological malingering. Behavioral Sciences and the Law, 2, 67-73.

Gough, H. (1947). Simulated patterns on the MMPI. Journal of Abnormal and Social Psychology, 42, 215-225.

Grisso, T. (1986a). Psychological assessment in legal context. In Curran, W., McGarry, A., & Shah, S. Forensic psychiatry and psychology: Perspectives and standards for interdisciplinary practice (pp. 103-128). Philadelphia: F.A. Davis.

Grisso, T. (1986b). Evaluating competencies: Forensic assessments and instructions. New York: Plenum.

Grisso, T. (1987). The economic and scientific future of forensic psychological assessment. American Psychologist, 42, 831-839.

Grossman, H. (Ed.). (1983). Terminology and classification manual of the American Association of Mental Retardation. Washington,D.C.: AAMR.

Grossman, L. & Wasyliw, O. (1988). A psychometric study of stereotypes: Assessment of malingering in a criminal forensic group. Journal of Personality Assessment, 52, 549-563.

Gudjonsson, G. & Shackleton, H. (1986). The pattern of scores on Raven's Matrices during "faking bad" and "non-

faking" performances. British Journal of Clinical Psychology, 25, 35-41.

Hartings, M. (1989). Case studies in detecting deception. Forensic Reports, 2, 223-233.

Hawk, G. & Cornell, D. (1989). MMPI profiles of malingerers diagnosed in pretrial forensic evaluations. Journal of Clinical Psychology, 45, 673-678.

Heaton, R.; Smith, H.; Lehman, R. & Vogt, A. (1978). Prospects for faking believable deficits on neuropsychological testing. Journal of Consulting and Clinical Psychology, 46, 892-900.

Hulett, A. (1941). Malingering - a study. Military Surgeon, 89, 129.

Hunt, W. & Older, H. (1943). Detection of malingering through psychometric tests. U.S. Naval Medical Bulletin, 41, 13-18.

Ironson, G. & Davis, G. (1979). Faking high or low creativity scores on the Adjective Check List. Journal of Creative Behavior, 13, 139-145.

Jahoda, M. & Warren, N. (1966). Attitudes: Selected readings. Baltimore: Penguin.

Jastak, S. & Wilkinson, G. (1984). Wide range achievement test - revised. Wilmington, DE: Jastak Associates.

Jensen, A. (1980). Bias in mental testing. New York: The Free Press.

Jung, C. (1910). The association reaction method. American Journal of Psychology, 21, 219-240.

Kaufman, A. & Kaufman, N. (1983). Kaufman Assessment Battery for Children. Circle Pines, MN: American Guidance Service.

Kilgore, B. (1982). Detecting the malingerer: A checklist. In: Resnick, P. (1984). The detection of malingered mental illness. Behavioral Sciences and the Law, 2, 21-38.

Kuvin, S. (1986). Psychiatric disability evaluations: Malingering, advocacy, bias. American Journal of Forensic Psychiatry, 7, 10-23.

Langfeld, H. (1921). Psychophysical symptoms of deception. Journal of Abnormal Psychology, 15, 319-328.

Larry P. v. Riles. 793 F.2d 969 (9th Cir. 1984).

Lees-Haley, P. (1984). Detecting the psychological malingerer. American Journal of Forensic Psychology, 2, 165-169.

Lees-Haley, P. (1986, Feb.). Personal injury malingering. For the Defense, no volume number, 28-31.

Lewis, D., Pincus, J., Bard, B., & Richardson, E.(1987). Neuropsychiatric, psychoeducational, and family characteristics of 14 juveniles condemned to death in the United States. American Journal of Child and Adolescent Psychiatry, 145, 584-589.

Lezak, M. (1983). Neuropsychological assessment (2nd ed.). New York: Oxford University Press.

Loo, R. & Wudel, P. (1979). Birth order and person variables. Psychological Reports, 45, 280.

Luckasson, R. (Ed.). (1992). Terminology and classification manual of the American Association of Mental Retardation. Washington, D.C.: AAMR.

Luckasson, R. & Keyes, D. (in progress). Defense assessment: What every defense attorney should know about

BEST COPY AVAILABLE

psychological assessment.

Marcus, E. (1986). Response to Dr. Kuvin's letter, "Call for common sense." American Journal of Forensic Psychiatry, 7, 24, 40, 58.

Marston, W. (1920). Reaction time symptoms of deception. Journal of Experimental Psychology, 3, 72-87.

McGarry, A. (1986) Forensic psychiatric reports: Selected clinical topics and models for report presentation. In Curran, W., McGarry, A., & Shah, S. Forensic psychiatry and psychology: Perspectives and standards for interdisciplinary practice (pp. 79-102). Philadelphia: F.A. Davis.

Menolascino, F. (Ed.). (1970). Psychiatric approaches to mental retardation. New York: Basic Books, Inc.

Menninger, K. (1963). The vital balance. New York: Viking.

Mercer, J. (1973). Labelling the mentally retarded. Berkeley, CA: University of California Press.

Meyer, R. (1983). Clinician's handbook: The psychopathology of adulthood and late adolescence. Boston: Allyn & Bacon.

Mickenberg, I. (1981). Competency to stand trial and the mentally retarded defendant: The need for a multi-disciplinary solution to a multi-disciplinary problem. California Western Law Review, 17, 365-402.

Miller, R. & Germain, E. (1988). The retrospective evaluation of competency to stand trial. International Journal of Law and Psychiatry, 11, 113-125.

Mitchell, R. & Thompson, N. (1986). Deception: perspectives on human and nonhuman deceit. Albany, NY: State University of New York Press.

Noble, J. & Conley, R. (1992). Toward an epidemiology of relevant attributes. In Conley, R., Luckasson, R., & Bouthilet, G. The criminal justice system and mental retardation: Defendants and victims (pp.17-54). Baltimore, MD: Brookes.

Ogloff, J. (1990). The admissibility of expert testimony regarding malingering and deception. Behavioral Sciences and the Law, 8, 27-43.

Pankratz, L. (1979). Symptom validity testing and symptom retraining: Procedures for the assessment and treatment of functional sensory deficits. Journal of Consulting and Clinical Psychology, 47, 409-410.

Pankratz, L. & Erickson, R. (1990). Two views of malingering. Clinical Neuropsychologist, 4, 379-389.

Pankratz, L., Fausti, S. & Peed, S. (1975). A forced-choice technique to evaluate deafness in the hysterical or malingering patient. Journal of Consulting and Clinical Psychology, 43, 421-422.

Parents in Action on Special Education (PASE) v. Hannon, 506 F. Supp. 831 (N.D. Ill. 1980).

Parwatikar, S., Holcomb, W. & Menninger, K. (1985). The detection of malingering amnesia in accused murderers. Bulletin of the American Academy of Psychiatry and the Law, 13, 97-103.

Patton, J.; Beirne-Smith, M.; Payne, J. (1990). Mental retardation (3rd Ed.). Columbus, OH: Merrill.

Penry v. Lynaugh, 109 S.Ct. 2934 (1989).

Perlin, M. (1987). The supreme court and the mentally disabled criminal defendant: Recent developments. Bulletin of the American Academy of Psychiatry and the Law, 15, 391-409.

Perske, R. (1991). Unequal justice? Nashville, TN: Abingdon Press.

President's Committee on Mental Retardation. (1970). The six-hour retarded child. Washington, D.C.: U.S. Government Printing Office.

Raven, J. (1958). Standard progressive matrices, sets A-E. London: Lewis & Co.

Resnick, P. (1984). The detection of malingered mental illness. Behavioral Sciences and the Law, 2, 21-38.

Ritson, B. & Forest, A. (1970). The simulation of psychosis: A contemporary presentation. British Journal of Medical Psychology, 43, 31-37.

Rogers, R. (1984). Towards an empirical model of malingering and deception. Behavioral Sciences and the Law, 2, 93-109.

Rogers, R. (Ed.). (1988). Clinical assessment of malingering and deception. New York: Guilford Press.

Rogers, R. (1990a). Models of feigned mental illness. Professional Psychology: Research and Practice, 21, 182-188.

Rogers, R. (Ed.). (1990b). Editor's note. Behavioral Sciences and the Law, 8, 1-2.

Rogers, R. (1990c). Development of a new classificatory model of malingering. Bulletin of the American Academy of Psychiatry and the Law, 18, 323-333.

Rogers, R., & Cavanaugh, J. (1983). "Nothing but the truth"... A reexamination of malingering. Journal of Psychiatry and the Law, 443-459.

Roman, D., Tuley, M., & Villaneuva, M. (1990). Evaluating MMPI validity in a forensic psychiatric population: Distinguishing between malingering and genuine psychopathology. Criminal Justice and Behavior, 17, 186-198.

Rosenham, D. (1972). On being sane in insane places. Science, 179, 250-258.

Sattler, J. (1988). Assessment of children (3rd ed.). San Diego: Author.

Scheerenberger, R. (1987). A history of mental retardation: A quarter century of promise. Baltimore, MD: Brookes.

Schretlen, D. (1986). Malingering: The use of a psychological test battery to detect two kinds of simulation. Ann Arbor, MI: University Microfilms International.

Schretlen, D. (1988). The use of psychological tests to identify malingered symptoms of mental disorder. Clinical Psychology Review, 8, 451-476.

Schretlen, D. & Arkowitz, H. (1990). A psychological test battery to detect prison inmates who simulate insanity or mental retardation. Behavioral Sciences and the Law, 8, 75-84.

Sierles, F. (1984). Correlates of malingering. Behavioral Sciences and the Law, 2, 113-118.

Shavelson, R. (1988). Statistical reasoning for the behavioral sciences (2nd ed.). Boston: Allyn & Bacon.

Spearman, C.(1927). The abilities of man. New York: Macmillan.

Spreen, O. & Benton, A. (1963). Simulation of mental deficiency on a visual memory test. American Journal of Mental Deficiency, 67, 909-913.

Swanson, D. (1984). Malingering and associated syndromes. Psychiatric Medicine, 2, 287-293.

BEST COPY AVAILABLE

Swanson, L. & Watson, B. (1989). Educational and psychological assessment of exceptional children (2nd ed.). Columbus, OH: Merrill.

Szasz, T. (1956). Malingering: "Diagnosis" or social condemnation? AMA Archives of Neurology and Psychiatry, 75, 432-443.

Tarjan, G., Wright, S., Eyman, R., & Keeran, D. (1973). Natural history of mental retardation: Some aspects of epidemiology. American Journal of Mental Deficiency, 77, 369-379.

Thorndike, E., Hagen, E., & Sattler, J. (1986a). Guide for administering and scoring the Stanford-Binet Intelligence Scale Fourth Edition. Chicago: Riverside.

Thorndike, E., Hagen, E., & Sattler, J. (1986b). Technical manual for the Stanford-Binet Intelligence Scale Fourth Edition. Chicago: Riverside.

Traven, S. & Protter, B. (1984). Malingering and malingering-like categories. Psychiatric Quarterly, 56., 189-197.

United States v. Hinckley 525 F. Supp. 1342 (1981).

Wachspress, M.; Berenberg, A.; & Jacobson, A. (1953). Simulation of psychosis. Psychiatric Quarterly, 27, 463-473.

Wainer, H. & Braun, H. (1988). Test validity. Hillsdale, NJ: Lawerence Erlbaum Associates.

Walters, G. (1988). Assessing dissimulation and denial on the MMPI in a sample of maximum security, male inmates. Journal of Personality Assessment, 52, 465-474.

Wasyliw, O., & Cavanaugh, J. (1989). Simulation of brain damage: Assessment and decision rules. Bulletin of the American Academy of Psychiatry and the Law, 17, 373-386.

Wasyliw, O.; Grossman, L.; Haywood, T.; & Cavanaugh, J. (1988). The detection of malingering in criminal forensic groups: MMPI validity scales. Journal of Personality Assessment, 52, 321-333.

Wechsler, D. (1981). Manual for the Wechsler Adult Intelligence Scale - Revised (WAIS-R). San Antonio: Psychological Corp.

Wertham, F. (1949). The show of violence. New York: Doubleday.

Wilbur, H. (1852). Third and final report of the experimental school for teaching and training idiotic children. Journal of Insanity, 9, 20-36.

Winick, B. (1985). Restructuring competency to stand trial. UCLA Law Review, 32, 921-984.

Yerkes, R. & Berry, B. (1909). The association reaction method of mental diagnosis. American Journal of Psychology, 20, 22-37.

Zigler, E., Balla, D., & Hodapp, R. (1984). On the definition and classification of mental retardation. American Journal of Mental Deficiency, 89, 215-230.

Ziskin, J. (1981). Coping with psychological and psychiatric testimony (3rd ed.). Venice, CA: Law and Psychology Press.

Ziskin, J. (1984). Malingering of psychological disorders. Behavioral Sciences and the Law, 2, 39-49.